LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 69
maggio 1974
anno VII
volume XII

La soia

È tra le voci piú importanti delle esportazioni degli Stati Uniti alla pari col frumento e poco dopo il mais. Ha perciò una funzione di primo piano nell'equilibrare la bilancia dei pagamenti americana

di Folke Dovring

Nel 1930 furono coltivati a soia 400 000 ettari di terreno negli Stati Uniti e il raccolto fu di circa 381 000 tonnellate. Lo scorso anno furono coltivati a soia oltre 22 milioni di ettari e la produzione complessiva viene stimata a piú di 40 milioni e 800 tonnellate. Ancor piú notevole è la crescente importanza della soia nelle esportazioni degli Stati Uniti. Fino al 1950 le esportazioni di soia erano trascurabili, mentre lo scorso anno raggiunsero un valore superiore a 3 miliardi di dollari, pari al 5 per cento circa del valore complessivo di tutte le esportazioni. Quest'anno la soia dovrebbe incidere nelle esportazioni degli Stati Uniti per un valore percentuale ancora piú elevato, diventando perciò un fattore determinante nella bilancia commerciale della nazione come fonte di valuta necessaria per il pagamento di importazioni essenziali come il petrolio greggio.

Il valore economico della soia risiede nel suo contenuto in proteine e grassi (estratti sotto forma di olio di soia). Benché il contenuto in grassi della soia (18-20 per cento) sia inferiore a quello dei semi della maggior parte delle altre piante oleifere, la produzione di grassi per ettaro è spesso superiore a quella delle altre piante perché la soia dà raccolti di semi molto maggiori. Il solo olio non consentirebbe però alla soia di competere con altre piante coltivate negli Stati Uniti; di importanza pari al contenuto in grassi è il tenore di proteine, che arriva fino al 40-50 per cento del peso dei semi, un rapporto insolitamente alto per qualsiasi tipo di pianta coltivata.

La soia è stata importata in America relativamente tardi. La pianta è originaria della Cina, ma (a differenza di molte altre piante) non si diffuse nel mondo nel periodo delle scoperte geografiche. Fino all'inizio del nostro secolo la soia veniva coltivata quasi per intero nell'Asia orientale. I semi di soia sono tra gli alimenti piú importanti nelle diete principalmente vegetariane della Cina, del Giappone e dell'Indonesia, ma pare abbiano contribuito assai poco, per lungo tempo, alla dieta, dominata dalla carne, dell'Occidente, dove furono usati soprattutto per fare salsa piccante a base di soia usata nei ristoranti cinesi e nella preparazione dei *ketchup* inglesi. Tentativi di coltivare la soia in Europa hanno avuto scarso successo a causa del clima, caratterizzato da siccità nella prima parte dell'estate. In Russia si è tentato di coltivare questa pianta già verso la fine dell'ottocento, ma qualche successo è stato ottenuto soltanto in epoca sovietica nelle regioni dell'Estremo Oriente.

Sia l'America settentrionale sia quella meridionale, con le loro grandi estensioni caratterizzate da un'alta piovosità estiva, offrivano possibilità climatiche migliori, e piú simili a quelle dell'Asia orientale, alla coltivazione di soia su grande scala. Anche le zone umide dell'Asia meridionale e dell'Africa tropicale hanno caratteristiche tali da essere potenzialmente paesi produttori di soia. La soia può essere coltivata con un sistema adeguato d'irrigazione in ogni paese in cui ci sia una estate calda.

La soia ebbe scarsa importanza negli Stati Uniti finché, durante la seconda guerra mondiale, non cominciò a scarseggiare il burro. In coincidenza con l'uso sempre piú esteso di olio di semi di soia nella margarina, i terreni coltivati a soia salirono a 4 milioni di ettari nel 1943, a 12 milioni di ettari nel 1964, a 16 milioni verso la fine degli anni sessanta. I piú di 22 milioni di ettari coltivati a soia lo scorso anno hanno portato questa coltura alla pari col frumento e poco dopo il mais.

La coltivazione della soia cominciò a diffondersi nella zona del mais. In anni recenti la coltura della soia è divenuta sempre piú importante nella valle del corso inferiore del Mississippi e negli Stati Uniti sud-orientali. Trenta stati hanno una produzione significativa di soia; i primi sette stati produttori di soia nel 1973 sono risultati, nell'ordine: l'Illinois, lo Jowa, l'Indiana, il Missouri, il Minnesota, l'Arkansas e l'Ohio. La coltura della soia non viene praticata in misura importante negli stati occidentali per le stesse ragioni climatiche per cui questa coltivazione non si è diffusa in Europa.

La soia è una pianta appartenente alle leguminose. Come altri membri della stessa famiglia, per esempio i piselli e il trifoglio, può ospitare sulle sue radici batteri fissatori di azoto e procurarsi pertanto da sé una parte dei fertilizzanti di cui ha bisogno. Il raccolto della soia è molto inferiore a quello del mais e del frumento nelle aree a elevata piovosità dove avviene la maggior parte della sua coltivazione. Dato però che il prezzo del prodotto è piú elevato, in anni recenti la soia ha superato il frumento come valore monetario complessivo della produzione. Con i prezzi pagati attualmente, il

raccolto della soia relativo al 1973 può rivaleggiare in valore monetario con quello del mais.

Tutto ciò ha fatto degli Stati Uniti il maggior produttore mondiale di soia (*si veda la figura a pagina 12 in alto*). Tre quarti circa dell'intera produzione mondiale di soia vengono coltivati negli Stati Uniti, i quali controllano quindi il grosso del commercio mondiale in questo campo. Il secondo paese produttore è la Cina, la cui produzione non è però cresciuta molto negli ultimi decenni; questo paese inoltre esporta solo una piccola parte del prodotto. L'unico altro paese esportatore importante è il Brasile, che negli ultimi anni ha molto sviluppato la coltivazione della soia ed esporta gran parte del raccolto. Gli altri principali produttori di soia, come l'URSS, il Giappone e l'Indonesia, producono ciascuno meno di un milione di tonnellate all'anno; il Giappone è costretto addirittura a importarne in grande quantità. Molti altri paesi, tra cui particolarmente il Messico e la Colombia, hanno cominciato a coltivare la soia a livello commerciale, ma la loro produzione non incide ancora in misura significativa sui mercati mondiali.

La crescente importanza economica della soia negli Stati Uniti ha influito profondamente sia sull'economia alimentare americana sia sul commercio internazionale. Un confronto fra le cifre relative alla produzione di burro e di margarina illustra la parte che la soia si è trovata a occupare nell'economia alimentare degli Stati Uniti. Nel 1940 il rapporto fra il burro e la margarina nei consumi interni era di 7: 1; oggi la margarina sta al burro nel rapporto di 2: 1. La maggior parte dell'accresciuta produzione di margarina è realizzata con l'olio di soia, al quale si devono ora tre quarti dell'intera produzione nazionale di margarina.

Eppure la margarina si aggiudica soltanto un quinto circa dell'olio di soia prodotto negli Stati Uniti. Con l'olio di soia si producono anche altri alimenti, come olio per condimento e grassi usati in pasticceria. L'olio di soia entra anche in numerosi prodotti non alimentari e nel commercio d'esportazione. Dall'olio di soia è ricavata oggi più della metà dei grassi commestibili prodotti negli Stati Uniti. In tutto il mondo la soia è diventata la fonte maggiore di grassi alimentari, superando il burro e anche qualsiasi altro tipo di olio di semi. La soia fornisce il 15 per cento circa dell'intera produzione mondiale di grassi commestibili, e circa un quarto di tutto il grasso commestibile presente nel commercio internazionale è costituito da olio di soia oppure da grasso estratto da semi di soia.

A causa del contenuto relativamente esiguo di grassi presenti nei semi di soia, la parte dei semi che rimane dopo l'estrazione del grasso vale un po' più dell'olio, anche se l'olio ha un prezzo più elevato per unità di peso. Una volta estratto l'olio, con quanto rimane dei semi di soia si producono principalmente farine per uso alimentare. Prescindendo dalle graminacee come il mais e il sorgo, i prodotti a base di soia costituiscono oggi circa la metà di tutte le sostanze commestibili in commercio negli Stati Uniti. La soia ha un'importanza ancora maggiore come costituente di mangimi per animali largamente esportati dagli Stati Uniti. Con l'aumento della produzione di carni di ogni tipo, e particolarmente di pollame, è della massima importanza la produzione di panello e di farina di soia. Le esportazioni di questi prodotti da parte degli Stati Uniti sono importanti anche per l'espansione zootecnica e quindi per la produzione di carne nell'Europa occidentale e (più recentemente) in Giappone.

Negli scambi con l'estero la soia è diventata, negli ultimi cinque anni, la voce più importante delle esportazioni agricole statunitensi, superando in valore monetario sia il frumento sia il mais. La svalutazione del dollaro all'inizio dello scorso anno, congiunta a una forte richiesta di prodotti agricoli, ha portato i prezzi di tali prodotti a livelli senza precedenti. Il raccolto di soia del 1973, una volta completato, si rivelò pari all'aspettativa di 41-44 milioni di tonnellate; se il 40 o il 50 per cento di tale produzione verrà esportata ai prezzi correnti, la soia potrebbe incidere per il 10 per cento sul valore monetario delle esportazioni degli Stati Uniti per il 1974.

L'alto livello attuale della domanda di soia americana potrebbe rivelarsi una condizione eccezionale temporanea. Quali sono le prospettive a lungo termine? La soia potrebbe assumere una posizione ancor più strategica nell'economia nazionale? La risposta dipenderà da due fattori: la tendenza a un maggiore rendimento per ettaro nella coltivazione della soia e la tendenza a un consumo più intensivo di proteine estratte dai semi di soia nell'alimentazione umana, sotto forma di « carne artificiale » o in altre forme.

La biologia della soia è stata ed è oggetto negli Stati Uniti di ricerche notevoli. Sono state sviluppate varietà di soia idonee alla coltivazione in varie parti del paese. I raccolti sono andati crescendo, lentamente ma in modo costante.

Altri paesi, oltre agli Stati Uniti, hanno beneficiato di queste ricerche, per esempio il Messico e l'India, i quali hanno importato recentemente dagli Stati Uniti varietà di soia che si sono rivelate appropriate alle condizioni climatiche locali. I risultati sono particolarmente promettenti in India, dove talune varietà americane hanno dato raccolti superiori a quelli che forniscono negli Stati Uniti. Le potenzialità della produzione di soia in India appaiono molto elevate.

Benché il rendimento della produzione della soia sia generalmente più alto di quello fornito da altre piante oleifere o da altre leguminose come piselli, fagioli e fave, i biologi che studiano la soia sono d'accordo sulla tesi che le potenzialità della coltura della soia siano lontane dall'aver raggiunto il loro tetto. È probabile che il rendimento della coltivazione della soia possa essere elevato fino a raggiungere un livello per ettaro doppio o triplo rispetto a quello attuale. Se tali risultati potranno essere raggiunti, saranno corrispondentemente maggiori le prospettive connesse alla produzione di cibi ad alto tenore proteico.

La produzione di cibi con proteine ricavate dalla soia per il consumo diretto da parte dell'uomo non è una cosa del futuro. Surrogati della carne e altri piatti a base di proteine ricavate dalla soia, che molte persone trovano gradevoli, sono già sul mercato. Una diffusa accettazione da parte dei consumatori può essere solo questione di tempo. A tal fine saranno importanti anche una promozione pubblicitaria di ricette a base di soia e una politica economica consapevole che favorisca la soia come cibo per l'uomo.

La produzione di surrogati della carne a base di soia si fonda su una tecnologia che ha compiuto buoni progressi in breve tempo. Le proteine della soia sono state disponibili sul mercato degli Stati Uniti a partire dal 1960; nel 1970 la produzione annua fu di circa 20 000 tonnellate. Le proteine estratte dalla soia ricevono una struttura fibrosa e il prodotto viene poi reso simile a qualsiasi tipo di carne, di cui gli viene conferito anche il sapore. Il nome usato per indicare le fibre è *textured vegetable protein*, ovvero proteina vegetale tessuta, o strutturata, abbreviato di solito in TVP. Secondo un regolamento emanato dal Ministero dell'agricoltura degli Stati Uniti, il materiale deve avere « un'integrità strutturale e una struttura identificabile, in modo che ciascuna unità resista all'idratazione e alla cottura e ad altri procedimenti usati nella preparazione di cibi per il consumo ». In altri termini, il prodotto deve conservare l'aspetto e il sapore della carne anche dopo essere bollito o fritto.

Una ricerca di mercato concernente un surrogato di prosciutto a base di farina di soia compiuta a Fort Wayne, nell'Indiana, nel 1970, rivelò una pronta accettazione del prodotto da parte del pubblico. Il prezzo, dopo la cottura (che nel prosciutto naturale dà luogo a una perdita di peso), è pari a circa la metà di quello del prosciutto naturale. Oggi i surrogati di carne sono disponibili al pubblico sotto due forme: bocconcini di prosciutto aromatizzato e colorato in modo da assomigliare al prodotto naturale, e proteine vegetali strutturate in fibre ma senza il sapore della carne. I bocconcini di prosciutto artificiali sono raccomandati come ingredienti di piatti in cui dovrebbe entrare il prosciutto naturale. Le proteine vegetali strutturate in fibre, che non hanno ricevuto il sapore della carne, vengono mescolate alla carne in hamburger o in altri piatti il cui ingrediente principale sia la carne tritata. Anche in questi casi il prodotto di soia costa circa la metà della carne vera.

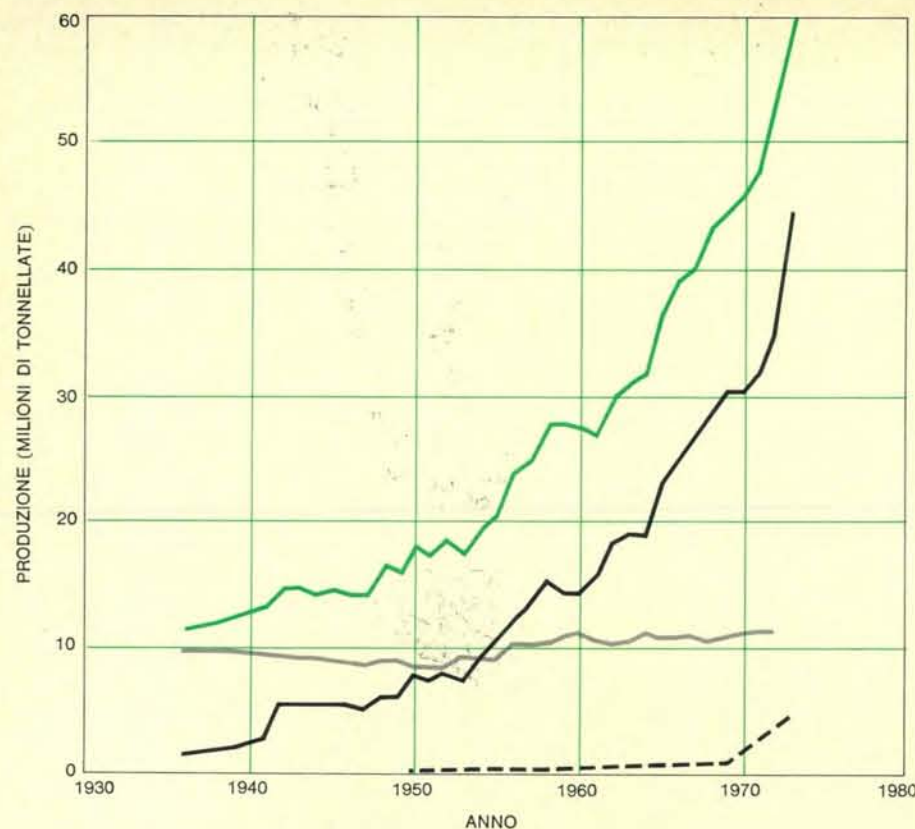
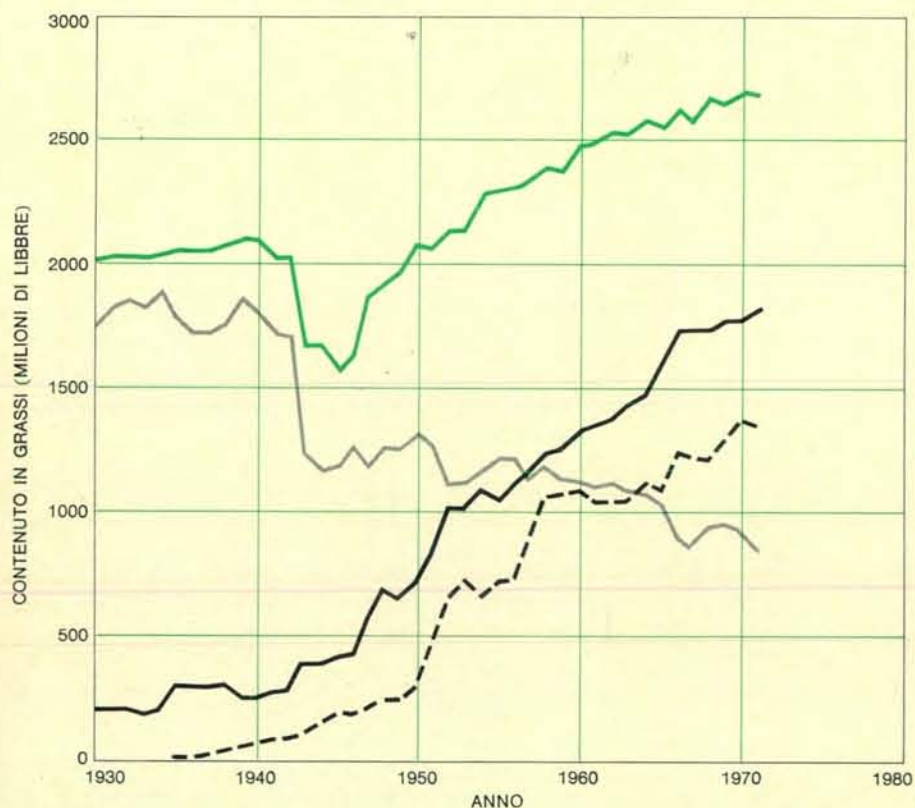
I produttori di questi surrogati della carne non hanno dato l'impressione di darsi molto da fare per accaparrarsi una grande fetta del mercato della carne. Questa prudenza si spiega in parte col fatto che taluni prodotti di soia messi sul mercato molti anni fa senza una valorizzazione adeguata hanno dato ai cibi a base di soia una cattiva fama. Può darsi anche che i produttori desiderino evitare conflitti con gli agricoltori e le industrie alimentari che lavorano la carne, alcuni dei quali sono membri delle stesse società o gruppi cui appartengono anche i produttori di proteine vegetali strutturate.

Quali che siano le ragioni dell'atteggiamento prudente dei produttori nei confronti dei surrogati della carne, rimane il fatto che il pubblico americano sta già consumando prodotti a base di soia in una misura superiore a quella indicata dalla produzione di proteine vegetali strutturate. Prodotti per la prima colazione, biscotti, vari tipi di pane, salsicce e molti altri cibi contengono farina di soia in forme meno costose rispetto alle proteine vegetali strutturate. Nel 1970 tali cibi impiegarono una quantità di proteine di soia 10 volte superiore a quella usata nella produzione di proteine vegetali strutturate. Taluni alimentaristi americani hanno realizzato recentemente una bevanda a base di soia che potrebbe sostituire il latte e hanno anche migliorato il sapore dei semi di soia per consentirne il consumo diretto nella loro forma naturale (per esempio in insalata), sotto forma di fiocchi in alimenti per bambini e in altre forme.

Gruppi religiosi e sanitari interessati a diete vegetali hanno sviluppato molti più prodotti e ricette a base di soia di quanti siano disponibili sul mercato generico. Ne consegue che una grande varietà di surrogati di carne possono essere trovati negli Stati Uniti in negozi di specialità. Chi desideri inoltre sperimentare nella propria cucina, può scoprire per i surrogati di carne un numero di usi molto maggiore rispetto a quelli suggeriti dai produttori. Per esempio, se il prodotto non aromatizzato viene mescolato con un surrogato di prosciutto fortemente aromatizzato, il miscuglio dà buoni risultati in ricette che richiedono carne tritata.

La proteina di soia è, dal punto di vista nutritivo, un po' meno completa della carne ma le poche carenze possono essere compensate facilmente usando altri vegetali, tra cui il mais. È perciò facilmente realizzabile una dieta completa senza prodotti animali. Numerosissimi altri surrogati della carne oltre a quelli attualmente disponibili potrebbero essere prodotti utilizzando la soia, compresi surrogati di vari tipi di carne senz'osso e di parti di pollo.

Il passaggio a una dieta composta in misura maggiore di vegetali presenterebbe per il pubblico vari vantaggi,

La figura illustra una pianta di soia, con le foglie disposte in modo da consentire la visione dei baccelli. Una pianta matura, alta 90 centimetri o più, avrebbe più fusti e foglie, e i baccelli si vedrebbero con difficoltà. In basso, un baccello maturo con due semi.

La produzione di soia negli ultimi quattro decenni: la produzione mondiale (*in colore*), degli Stati Uniti (*in nero*), della Cina (*in grigio*) e del Brasile (*tratteggiata*).

Il mutare del rapporto fra il consumo di burro e quello di margarina negli Stati Uniti a partire dal 1930 spiega il grande aumento dell'impiego di olio di soia nella produzione di margarina (*curva tratteggiata*). Il consumo totale di burro e di margarina è cresciuto (*in colore*), ma la margarina (*in nero*) occupa oggi una parte di mercato molto maggiore di quella del burro (*in grigio*), capovolgendo il rapporto precedente.

tra cui una riduzione del costo della vita e dell'ingestione di grasso. Esso diminuirebbe anche la pressione dell'attività economica nei confronti dell'ambiente, poiché l'attuale crisi energetica è sintomatica dell'incombere di una scarsità di risorse che potrebbe essere alleviata in misura notevole se una parte minore dei terreni agricoli fosse usata per nutrire animali domestici.

La produzione di prosciutto naturale, per esempio, richiede un'estensione di colture 10 volte superiore a quella richiesta per produrre la stessa quantità di surrogato di prosciutto tratto dalla soia; la produzione di carni bovine richiede un'estensione di terreni da 15 a 20 volte superiore a quella necessaria per la produzione di un'ugual quantità di soia. Poiché inoltre la soia può produrre, grazie ai batteri che ospita fra le proprie radici, una parte dei fertilizzanti azotati di cui abbisogna, attinge in minore quantità alle industrie produttrici di fertilizzanti, le quali fanno un uso intensivo di energia. Diminuirebbe pertanto anche la pressione ecologica legata ad alti livelli di applicazione di fertilizzanti azotati, poiché minori quantità di azoto verrebbero trasportate via dall'acqua del sottosuolo e dai corsi d'acqua.

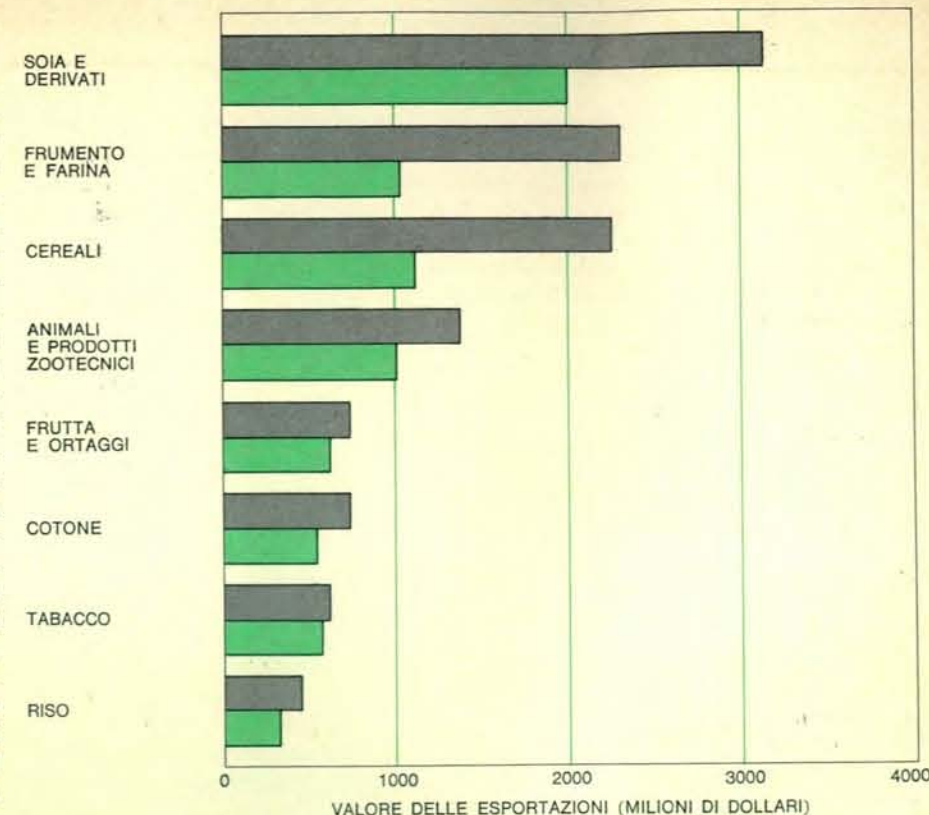
Per paesi aventi un basso reddito pro capite, cibi basati su proteine vegetali offrono vantaggi ancora maggiori. Benché attualmente nei paesi piú poveri la soia sia consumata per lo piú direttamente (o sotto forme come quelle di additivi della farina e di « latte di soia »), a lungo termine i surrogati della carne potrebbero diventare molto importanti nello sviluppo economico. L'esperienza dimostra che, in coincidenza con l'aumento del reddito pro capite, cresce la richiesta di carne. Tale richiesta aumenta con maggiore rapidità a taluni livelli di reddito intermedi. « L'inflazione si nutre di carne rossa » è una frase sorta in Francia fra il 1950 e il 1960, quando il paese si trovava a un tale livello medio di reddito. L'espressione ha oggi un significato particolarmente vero in paesi come l'Argentina e il Messico. Prima o poi, se tutto procede regolarmente, il problema si porrà nei paesi attualmente con un reddito basso. Per tali paesi sarebbe economicamente molto piú semplice progettare grandi approvvigionamenti di proteine vegetali che non cercare di impiantare una produzione di carne su una scala tale da soddisfare le aspirazioni della popolazione, poiché, le proteine vegetali sono molto piú economiche. Di fatto la situazione alimentare del mondo intero trarrebbe

un grande beneficio da un maggiore consumo di proteine vegetali.

Per gli Stati Uniti la questione delle priorità agricole è strettamente legata a problemi di commercio internazionale. Negli ultimi anni il paese si è trovato in una situazione di sempre maggiore dipendenza da importazioni di petrolio, particolarmente dalle nazioni arabe del Medio Oriente e dell'Africa settentrionale. Prescindendo dall'inopportunità politica di dipendere da fonti cosí lontane, la nazione deve far fronte all'incidenza delle importazioni di petrolio sulla bilancia dei pagamenti. La situazione della bilancia dei pagamenti è abbastanza negativa ed è andata peggiorando fino alla prima parte del 1973. Un modesto surplus di pagamenti negli ultimi mesi ha riscattato soltanto per una frazione molto piccola le quantità enormi di dollari tesaurizzati in paesi stranieri. Oggi si ha l'impressione che le esportazioni agricole, con la soia al primo posto, possano recar sollievo agli Stati Uniti, almeno per qualche tempo. Il problema è quello di sapere fino a quando potranno continuare a farlo. La risposta si trova, tra l'altro, nelle tendenze che si avranno nel consumo di carne e di energia.

Entrambe le tendenze hanno presentato un'ascesa notevole e nei primi anni del decennio scorso sono salite a ritmi superiori a quelli dei vari decenni precedenti. Questi due fenomeni sono chiari indizi dell'aumento del livello di vita. Dalle cifre risulta chiaro che la popolazione degli Stati Uniti in generale dispone ora di un reddito maggiore che nell'ultimo decennio e che il miglioramento è stato piú rapido che in decenni precedenti.

Questi sviluppi sarebbero tutti per il meglio se non fosse per il problema delle risorse implicito nell'aumento delle cifre corrispondenti al consumo di carne e di energia. Il problema è reso piú grave dal fatto che gran parte del crescente consumo di carne è a carico dei bovini. Attualmente gli Stati Uniti stanno pagando le proprie importazioni (di carne oltre che di petrolio) con l'esportazione di soia, frumento, mais e altri prodotti agricoli non lavorati. Se la tendenza al consumo di carne dovesse continuare a salire, una parte maggiore della produzione agricola della nazione dovrebbe essere utilizzata nell'alimentazione del bestiame e ne sarebbe quindi disponibile per l'esportazione una parte minore. Continuando a salire il consumo di energia, ci si deve attendere anche un aumento nelle importazioni di petrolio.

In altri termini, le tendenze all'aumento che si riscontrano nel consumo

Le esportazioni agricole degli Stati Uniti sono dominate dalla soia e dai suoi derivati. Le barrette rappresentano le esportazioni di beni principali negli anni terminanti il 30 giugno 1972 (*in colore*) e il 30 giugno 1973 (*in grigio*). La soia copre attualmente piú del 5 per cento del valore complessivo dell'intera esportazione degli Stati Uniti.

della carne e in quello dell'energia sono su una rotta di collisione. Se continueranno a salire entrambe, gli Stati Uniti verranno a trovarsi in crescenti difficoltà. Per esprimere la cosa in termini piú precisi, può darsi che gli americani vengano a trovarsi dinanzi all'alternativa di dover scegliere fra bistecche e viaggi di piacere. Una svolta verso i trasporti di massa e verso altri mezzi per economizzare i consumi di energia potrebbe ridurre il fabbisogno e quindi le importazioni di petrolio. Inoltre, uno spostamento dell'interesse del pubblico verso surrogati della carne e verso cibi a elevato tenore di proteine di origine vegetale – con in testa la soia – potrebbe ridurre il fabbisogno interno di mangimi per il bestiame e consentire maggiori esportazioni di prodotti agricoli con i quali pagare il petrolio e altre importazioni essenziali, mantenendo in equilibrio la bilancia dei pagamenti.

I prodotti agricoli hanno sempre avuto, storicamente, una grande incidenza nelle esportazioni degli Stati Uniti. Anche oggi il paese ha il piú grande potenziale agricolo del mondo. (In futuro potrebbe rivaleggiare con gli Stati Uniti il Brasile, ma la cosa rimane da vedere.) È ovvio che gli Stati Uniti sfruttino la loro posizione di vantaggio in questo campo pagando le loro importazioni con i proventi delle loro esportazioni agricole.

Nei decenni recenti, però, le esportazioni agricole della nazione hanno avuto una parte relativamente minore mentre sono cresciute le esportazioni dei prodotti industriali, grazie alla posizione di vantaggio delle industrie americane sul piano tecnologico e organizzativo nei confronti delle industrie di altri paesi. Tra le esportazioni industriali c'erano molti prodotti tecnici.

Recentemente il vantaggio industriale degli Stati Uniti è declinato in qualche misura perché molti paesi hanno imparato la lezione e sono stati in grado di produrre a loro volta nel campo tecnologico. Questo fatto era probabilmente inevitabile; un vantaggio relativo fondato sulla tecnologia tende a durar poco in un periodo di rapidi progressi tecnologici. Piú sicuro è un vantaggio relativo fondato sulle risorse naturali; è ovvio perciò che gli Stati Uniti debbano puntare di nuovo sulla base agricola per rafforzare la loro posizione nel commercio con l'estero. Questa logica è ancor piú stringente ora che la domanda mondiale di prodotti agricoli degli Stati Uniti è cresciuta, in parte a causa dell'ascesa dei livelli di

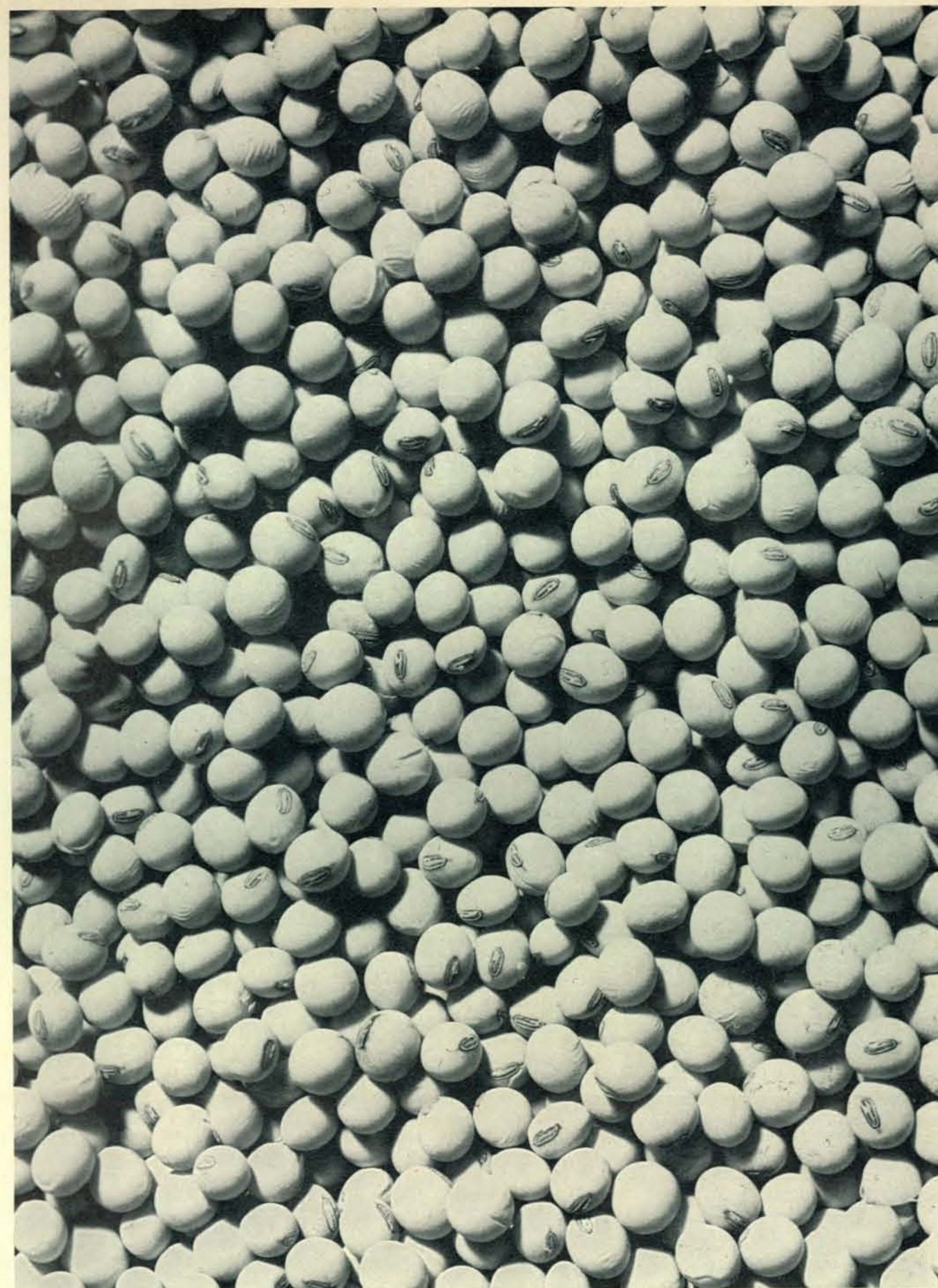
La coltura della soia in Italia

L'articolo evidenzia l'enorme importanza economica che, nel volgere di pochi anni, la coltura di questa leguminosa ha assunto negli Stati Uniti d'America, che oggi, superata anche la Cina, sono nel mondo il piú forte paese produttore ed esportatore di seme di soia. Nel 1972 gli USA hanno infatti ricavato dalle loro colture, con una resa media unitaria nazionale fra le piú elevate che si riscontrino, oltre il 65% dell'intera produzione mondiale. Ben diversa è la situazione in Europa, dove la soia è coltivata su modeste superfici soltanto in pochissimi paesi. Primeggia la Romania, che nel 1972 vi ha destinato 150 000 ettari.
Fra i paesi europei produttori di soia, figura, secondo le statistiche nazionali e internazionali, anche l'Italia, dove, introdotta oltre un secolo fa, questa coltura non si è mai consistentemente diffusa, pur avendo sempre suscitato grande interesse negli operatori, nei tecnici e negli studiosi. Cosí, nell'ultimo quarantennio, vi ha occupato superfici oscillanti da qualche decina di ettari a un massimo di 2-3000 nel 1947-1948. Nell'ultimo biennio ha interessato, quasi esclusivamente in alcune province del Veneto, una sessantina di ettari, per quanto questo dato, desunto dalla statistica ufficiale, appaia rilevato forse piú in eccesso che in difetto rispetto alla situazione reale.
L'importanza economica che in campo mondiale riveste la soia deriva dal fatto che dal seme, ricco in protidi (38-40%), in lipidi (18-20%) e in fosfatidi, si ricavano prodotti alimentari e industriali sempre piú richiesti da tutti i mercati, quali olio per uso alimentare e industriale, mangimi ad alto contenuto proteico largamente utilizzati in campo zootecnico, materie prime per l'industria dolciaria e per altre industrie minori. I protidi della soia, inoltre, sono caratterizzati dal contenere in elevata proporzione gli amminoacidi essenziali per la crescita, quali lisina (6%), leucina, arginina, isoleucina, valina, acido aspartico e acido glutammico. È quindi oggi una delle colture agrarie che forniscono, in via diretta e indiretta, uno dei piú alti contributi al rifornimento di alimenti protidici e lipidici di cui l'umanità manifesta un crescente bisogno.
In relazione a questa sua grande importanza economica, non può quindi, a prima vista, non meravigliare che la coltura della soia in Italia sia ancor oggi pressoché inesistente, mentre in altri paesi occidentali e orientali, situati a pari latitudine e caratterizzati da condizioni ambientali non dissimili dalle nostre, venga invece attuata con largo profitto e su superfici sempre piú vaste. Per soddisfare i fabbisogni di soia e dei prodotti derivati, siamo costretti a onerose importazioni. Nel biennio 1971-1972, abbiamo acquistato all'estero seme, olio e panello di soia per un valore medio annuo complessivo di 160 milioni di dollari.
I motivi che fino a oggi non hanno consentito una adeguata diffusione di questa coltura nel nostro paese, sono stati, nel passato, essenzialmente di ordine tecnico; in epoca piú recente, di carattere prevalentemente economico. Alcuni decenni fa vi era infatti scarsa disponibilità di varietà precoci e produttive e dotate di caratteristiche foto e termo-stadiali rispondenti al nostro ambiente di coltivazione; presentava difficoltà – non disponendosi commercialmente delle apposite colture batteriche (nitragine) – l'esecuzione della « assoiatura », pratica con la quale si inocula nel terreno il ceppo specifico del *Rhizobium leguminosarum*, che è assente ove questa coltura mai sia stata attuata, e che determina, per simbiosi con l'apparato radicale, la fissazione dell'azoto atmosferico da parte della pianta; vi era, infine, da risolvere una vasta problematica inerente la tecnica colturale (investimento unitario, concimazioni, ecc.). L'imponente attività di ricerca condotta negli ultimi decenni in molti paesi ha ormai consentito all'agricoltore, sia di disporre di un'ampia gamma di varietà adatte ai piú diversi ambienti, sia di attuare la coltivazione con la tecnica piú razionale.
Il problema economico, rappresentato soprattutto dalla scarsa remuneratività dei prezzi di vendita fortemente compromessa dalla concorrenza del prodotto proveniente dagli USA, è quello che si è fatto in seguito maggiormente sentire. Nel continente americano la soia è coltura estensiva e integralmente meccanizzata, per cui i costi di produzione risultano assai piú bassi che nel nostro paese, dove, fra l'altro, colture agronomicamente similari sono assai piú remunerative. Oggi, per la situazione determinatasi a seguito delle limitazioni alla esportazione della soia adottate dal governo americano e alle difficoltà in cui si trovano i grandi paesi produttori per soddisfare la domanda mondiale vertiginosamente crescente, il problema della competitività di questa coltura si presenta sotto un profilo migliore, sí che verso la soia si sta risvegliando l'interesse degli agricoltori italiani. (*R.F.*)

vita in Europa, Giappone e altrove e in parte perché molti paesi hanno maggiori difficoltà degli Stati Uniti nell'aumentare la loro produzione agricola. Le prospettive di scambi commerciali fra gli Stati Uniti e l'URSS si fondano essenzialmente sulla superiorità degli Stati Uniti nel campo dell'agricoltura e sulle grandi risorse minerarie dell'URSS.

Gli Stati Uniti recupereranno piú facilmente la loro posizione di fornitori di sostanze alimentari per gran parte del mondo se si riuscirà a diminuire o a bloccare all'interno la tendenza a dedicare una parte crescente dei terreni agricoli alla produzione di cibo per gli animali domestici in vista di un crescente consumo di carne. Le esperienze recenti sono significative: se il burro, sul mercato avesse continuato a prevalere sulla margarina, ora disporremmo di meno cereali per l'esportazione. In futuro, uno spostamento verso i surrogati della carne a base di proteine vegetali, e particolarmente di quelle della soia, che sono le piú economiche e abbondanti, potrebbe contribuire a rendere disponibili nuove estensioni di terreno agricolo per colture da esportare.

I nuovi orientamenti nella produzione, nel consumo e nel commercio impliciti nell'argomentazione qui esposta non possono aver luogo in modo brusco. Ciò che importa è guardare al futuro dei problemi nazionali ed essere consapevoli delle probabili conseguenze di linee d'azione alternative. Se la previsione fosse stata applicata alle prospettive a lungo termine dell'economia del decennio 1950-1960, per esempio, non avremmo avuto probabilmente né l'attuale crisi energetica né la sfavorevole bilancia dei pagamenti nel commercio con l'estero.

La superiorità della soia sia a fini interni sia per l'esportazione deriva dalla grande varietà dei prodotti che se ne possono ricavare. Una caratteristica generale importante di un'economia moderna è la crescente versatilità delle materie prime. Questa caratteristica, insieme all'aumento della produttività e del reddito pro capite, è tra le fonti della crescente libertà di scelta che fa dell'ambiente socioeconomico una « società opzionale », ossia una società in cui la gamma delle scelte che vengono offerte sia agli individui sia alla società nel suo complesso è piú vasta di quanto sia mai stata in passato e tende ancora ad ampliarsi. La soia ha da questo punto di vista una grande importanza e rappresenta quindi un bene eminentemente moderno. I biologi e gli alimentaristi che si occupano della soia hanno perciò una parte decisiva nel futuro economico degli Stati Uniti.

I semi di soia hanno forma pressappoco sferica e presentano spesso chiazze scure caratteristiche. Essi hanno approssimativamente la grandezza dei piselli. La produzione di soia del 1973 negli Stati Uniti è valutata tra i 41 e i 44 milioni di tonnellate.

I dispositivi ad accoppiamento di carica

Frutto di un nuovo concetto nell'elettronica dei semiconduttori, hanno un grande potenziale applicativo come sensori di immagine e come elementi di immagazzinamento di informazioni nella memoria dei calcolatori

di Gilbert F. Amelio

Negli scorsi quattro anni si è diffuso, tra i fisici dello stato solido, un interesse crescente nei confronti di un nuovo concetto nella elettronica dei semiconduttori che potrebbe un giorno influenzare profondamente la nostra vita, come oggi accade per il transistore. Si tratta del concetto di accoppiamento di carica che si traduce praticamente nel dispositivo ad accoppiamento di carica.

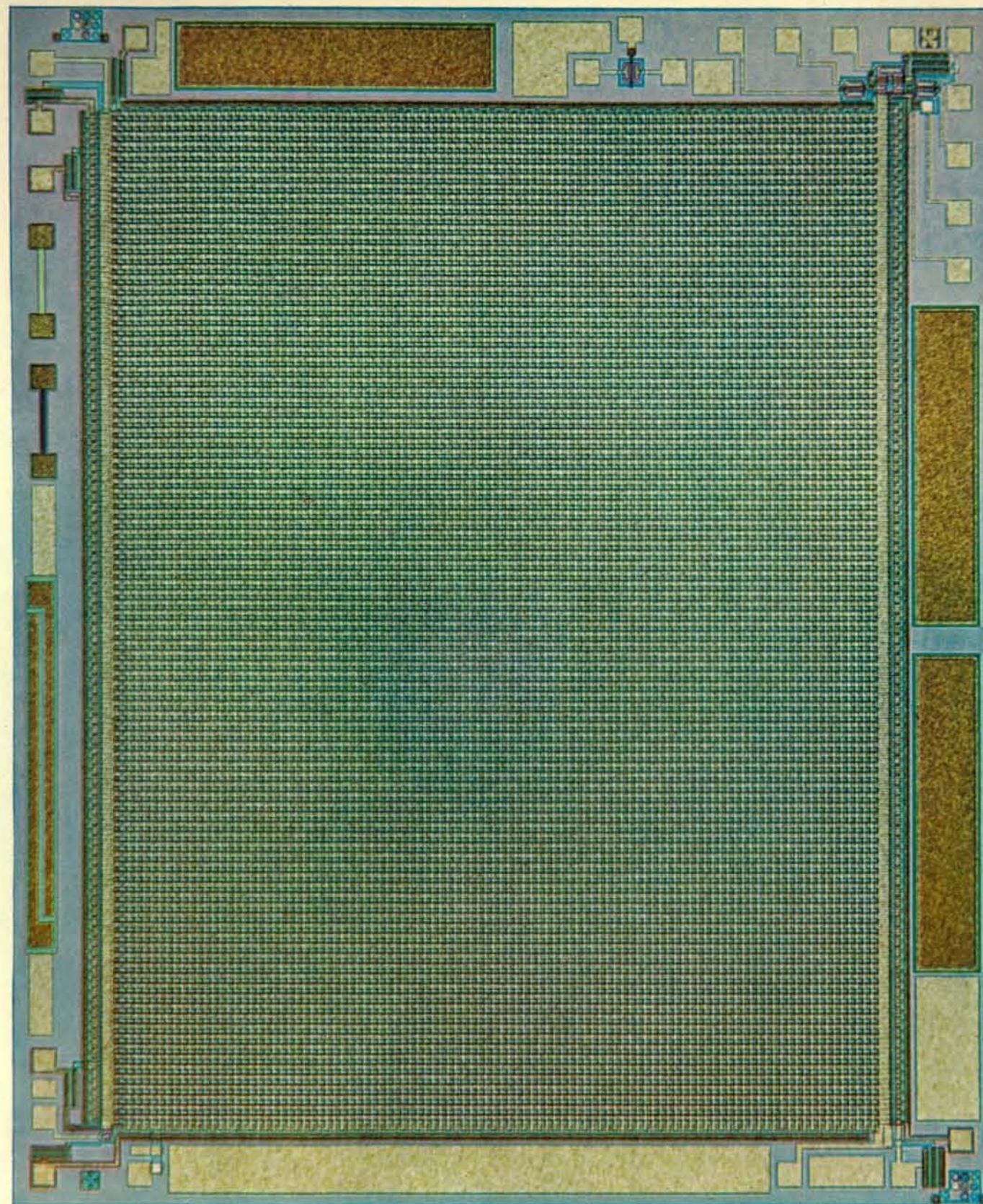
Come il transistore, il dispositivo ad accoppiamento di carica trova la sua base teorica nell'elettronica dei semiconduttori, ed è perciò soggetto alle medesime leggi fisiche che governano il funzionamento e la fabbricazione del transistore. Ma, a questo punto, ogni analogia finisce. Pur sfruttando in gran parte la stessa matrice tecnologica del suo diffuso predecessore, il dispositivo ad accoppiamento di carica è un elemento funzionale che agisce sulla trasformazione dell'informazione e non, come il transistore, un elemento attivo che agisce sulla modulazione delle correnti elettriche. La tecnologia del transistore ha reso possibile la realizzazione di componenti di memoria per calcolatori che contengono migliaia di elementi su una sola piastrina di silicio; il principio di accoppiamento di carica permetterà di produrre componenti di memoria delle stesse dimensioni contenenti decine o anche centinaia di migliaia di elementi di memoria per piastrina a circa lo stesso costo.

Immagine al microscopio elettronico a scansione dell'area in prossimità dei contatti di uscita del reticolo di un fotosensore ad accoppiamento di carica. Ogni elemento col relativo circuito rivelatore occupa un millesimo di millimetro quadrato.

Ma che cosa si intende per accoppiamento di carica? Con l'espressione accoppiamento di carica si definisce lo spostamento globale di tutta la carica elettrica mobile immagazzinata in un elemento di memoria a semiconduttore in un analogo elemento di memoria adiacente mediante un intervento esterno di modifica dei potenziali. La quantità di carica contenuta in questo « pacchetto » mobile può variare ampiamente, in relazione alle tensioni applicate e alla capacità elettrica dell'elemento di memoria. La carica elettrica posseduta da ogni pacchetto può rappresentare l'informazione.

Una semplice rappresentazione del funzionamento del dispositivo ad accoppiamento di carica può essere fornita dal seguente modello meccanico. Si immagini un meccanismo costituito da una serie di tre pistoni collegati a un albero a gomito dalle relative bielle (*si veda la figura in alto alle pagine 18 e 19*). Sulla parte superiore di uno o piú pistoni è posto del fluido. Si può facilmente constatare che ruotando l'albero a gomiti in verso orario si provoca lo spostamento del fluido verso destra, mentre, ruotando l'albero in senso antiorario, il fluido si sposterà verso sinistra. Poiché la ripetizione di questo schema richiede tre pistoni, questa disposizione viene chiamata sistema a tre fasi. Se invece si vuole spostare il fluido in una sola direzione, si può immaginare un sistema a due fasi, costituito da due pistoni a profilo asimmetrico (*si veda la figura in basso alle pagine 18 e 19*). Indipendentemente dal verso di rotazione dell'albero, il fluido avanzerà in questo caso sempre verso destra.

Concettualmente analoghi sono i dispositivi ad accoppiamento di carica, solitamente costruiti con silicio (*si vedano le figure a pagina 20*). Essi sono formati da un substrato di silicio di tipo *p* (in cui i portatori di segnale sono solitamente costituiti dagli elettroni) sul quale è diffuso uno strato isolante di biossido di silicio. Un reticolo di elettrodi conduttori viene depositato a sua volta sulla superficie isolante. Opportuni collegamenti fra gli elettrodi permettono di realizzare sia il funzionamento a due fasi sia quello a tre fasi. All'interno del semiconduttore, sotto lo strato isolante, si può alterare selettivamente la conducibilità del silicio in modo da formare zone di tipo *n* (in cui le lacune elettroniche sono i portatori di segnale). L'analogia col modello meccanico precedente è evidente se si suppone che il fluido rappresenti un accumulo di elettroni, i pistoni l'energia potenziale associata alle tensioni applicate agli elettrodi, e l'albero a gomiti e le bielle le tensioni pilota e la loro temporizzazione.

Quando un'opportuna forma d'onda periodica, chiamata tensione di *clock* (temporizzazione), viene applicata agli elettrodi, alcuni elettroni presso ciascun elettrodo formeranno un pacchetto discreto di carica spostando un elemento ad accoppiamento di carica, o cella elementare, verso destra ad ogni periodo del temporizzatore. Pertanto, i pacchetti di carica elettronica si spostano verso destra in conseguenza del continuo spostamento laterale del « pozzo

Il sensore di immagini ad accoppiamento di carica, rappresentato in questa fotografia realizzata dall'autore presso il laboratorio di ricerca e sviluppo della Fairchild Camera and Instrument Co., consta di 10 000 elementi fotosensibili disposti su una griglia rettangolare di 100 × 100 elementi in una piastrina di silicio di 4 × 5 millimetri. Le sagome brune e chiare lungo i lati sono quelle dei contatti metallici degli elettrodi. Il dispositivo è stato costruito per essere utilizzato in una telecamera sperimentale basata sul principio dell'accoppiamento di carica, che consiste nel trasferimento complessivo di tutta la carica elettrica mobile accumulata in un elemento semiconduttore in un altro elemento semiconduttore adiacente, ottenuto modificando dall'esterno le loro tensioni. Pur essendo i sensori di immagine ad accoppiamento di carica ancora a uno stadio sperimentale, l'autore afferma che essi « indicano già una promettente via per giungere alla realizzazione di efficaci telecamere ». Nella microfotografia della pagina a fronte è riprodotto un fortissimo ingrandimento di una porzione del reticolo fotosensibile.

Questo modello meccanico rappresenta il funzionamento di un dispositivo ad accoppiamento di carica. Il meccanismo è costituito da una serie di tre pistoni alternativi collegati a un albero a gomiti dalle relative bielle. Sulla sommità di uno o piú di essi è posto del fluido (*in colore*). Ruotando l'albero a gomiti in senso orario, come viene indicato nella figura, si provoca lo spostamento del fluido verso destra. Se invece l'albero a gomiti viene fatto ruotare in senso antiorario, il fluido si sposterà verso sinistra. Questo particolare schema costruttivo, che necessita di tre pistoni per ripetere il ciclo, si chiama sistema a tre fasi. Un dispositivo analogo ad accoppiamento di carica, cioè un sistema a tre fasi, è mostrato nella figura in alto di pagina 20.

di potenziale » locale in cui vengono a trovarsi: essi, cioè, seguono una traiettoria di caduta continua.

La creazione del « pozzo di potenziale » necessario nel substrato semiconduttore richiede alcuni chiarimenti perché costituisce il punto chiave nella teoria dell'accoppiamento di carica. In questo contesto si intende per « pozzo di potenziale » un volume localizzato nella piastrina di silicio, che esercita un'attrazione nei confronti degli elettroni; in altri termini, esso rappresenta il sito piú positivo in un certo intorno e costituisce perciò la naturale destinazione degli elettroni. Nell'elemento di memoria ad accoppiamento di carica i « pozzi di potenziale » sono formati dall'interazione di regioni di silicio a diverso tipo di conducibilità (*si veda la figura a pagina 21*) Tale interazione forma un pozzo tanto piú profondo quanto piú elevata è la tensione di *clock*. Tutti gli elettroni contenuti nel pozzo si sposteranno insieme alle tensioni di *clock*.

Ora, se due o piú pozzi aventi profondità diversa sono situati l'uno accanto all'altro, essi si sovrapporranno cosicché la carica può essere « accoppiata », o trasferita, da un elemento di memoria al successivo ogni volta che la profondità del pozzo subisce una variazione per effetto delle tensioni di *clock*. In questo modo, le tensioni di *clock* regolate dall'esterno agendo sugli elettrodi guidano il moto dei pacchetti di elettroni lungo un solco di energia potenziale, definito canale. Tale modalità di trasferimento di elettroni costituisce l'essenza dell'accoppiamento di carica.

Il fenomeno dell'accoppiamento di carica da solo non permette la costruzione di un dispositivo di utilità pratica. Un simile dispositivo deve essere in grado di introdurre nella struttura gli elettroni necessari al funzionamento e possedere, in qualche punto del canale, un mezzo capace di rivelare la quantità di carica contenuta in un pacchetto. Quindi, perché una struttura possa essere classificata come dispositivo ad accoppiamento di carica, dovrà possedere tre stadi: di ingresso, di accoppiamento di carica e di uscita.

Come esempio di un dispositivo completo ad accoppiamento di carica, consideriamo un « registro di scorrimento » formato da otto elementi a tre fasi, un diodo e « porta » di ingresso, un diodo e « porta » di uscita (*si veda la figura a pagina 22*). Tale struttura è effettivamente molto simile al primo dispositivo ad accoppiamento di carica costruito. Il segnale che si vuol fare entrare nel dispositivo è collegato al diodo di ingresso, che funge da sorgente di elettroni. Se la « porta » di ingresso viene mantenuta a basso potenziale, gli elettroni del segnale non possono entrare nel canale. Per introdurre un pacchetto di elettroni nel dispositivo è necessario attendere finché gli elettrodi della prima fase sono in condizione di potenziale elevato e quindi « aprire » la « porta » di ingresso elevandone il potenziale. Gli elettroni riempiono il « pozzo di potenziale » finché il livello di energia degli elettroni del pozzo non uguaglia quello degli elettroni della sorgente. La tensione alla « porta » di ingresso viene ora abbassata per isolare la sorgente, e il pacchetto di carica formatosi è pronto per essere trasferito lungo il canale. Nella fase di rivelazione, la carica viene semplicemente trasferita a uno « scarico », rappresentato da un diodo di uscita, dove compare come corrente in qualche circuito esterno. Questo semplice dispositivo ad accoppiamento di carica compie le funzioni di « registro di scorrimento » a otto bit, un dispositivo della memoria dei calcolatori.

Il funzionamento reale dei dispositivi ad accoppiamento di carica avviene nel modo descritto teoricamente, con una sola limitazione: in realtà non tutti gli elettroni del pacchetto avanzano a ogni trasferimento, ma una carica residua forma un pacchetto di coda. Il valore di questa « perdita di trasferimento di carica » è funzione delle caratteristiche di progetto del dispositivo e della frequenza di funzionamento. La perdita di trasferimento impone un limite alla velocità e al numero di trasferimento nel funzionamento pratico di un dispositivo ad accoppiamento di carica per effetto dell'attenuazione subita dal pacchetto mentre si sposta da una regione all'altra.

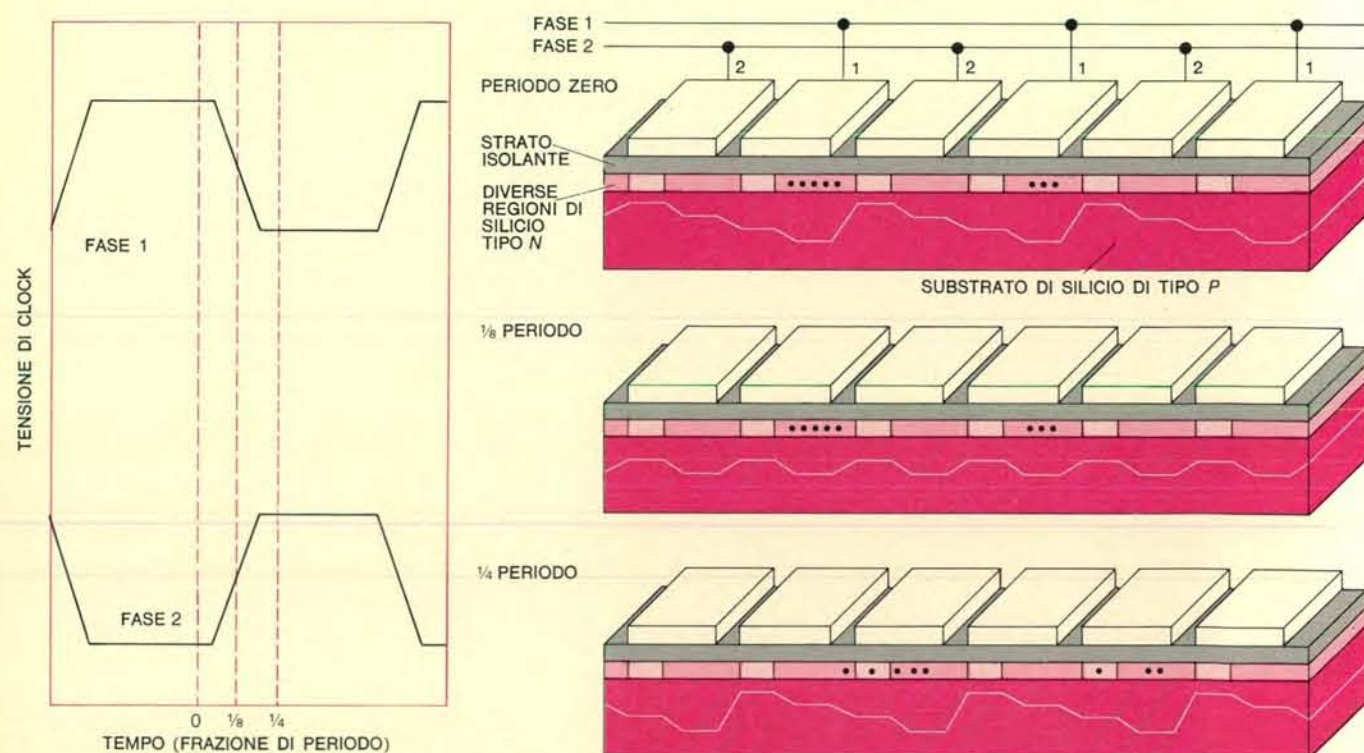
La perdita di trasferimento di carica è dovuta a due cause: gli elettroni possono venire arrestati da aree a minore energia potenziale (corrispondenti a cavità o sporgenze sulla faccia superiore del pistone nel modello meccanico); oppure la frequenza di funzionamento può risultare cosí alta da non permettere a tutti gli elettroni di seguire i « pozzi di potenziale ». Il primo problema è direttamente influenzato dai particolari costruttivi del singolo dispositivo ad accoppiamento di carica. Questo aspetto è stato oggetto di ricerche approfondite e recenti sviluppi tecnologici hanno ridotto significativamente l'entità del problema. La perdita legata alla velocità di spostamento ha invece origini piú profonde e richiede un ulteriore chiarimento.

Gli elettroni vengono indotti a spostarsi verso una regione adiacente a minore energia, cioè in un pozzo di potenziale piú profondo, dalla combinazione di tre fattori: forze autoindotte, forze legate al campo e forze di origine termica. Lo spostamento autoindotto dipende dal fatto che un pacchetto di elettroni (o di altre particelle simili) ad alta densità tende ad allargarsi rapidamente se viene rimossa la forza che lo mantiene assieme, come accade per esempio quando viene variata la tensione di *clock*. Questo tipo di forza prevale durante i primi stadi del trasferimento di carica. Lo spostamento legato al campo interviene se nel progetto della struttura esistono campi elettrici che facilitano il moto dell'elettrone nella direzione voluta. Questo equivale, nell'analogia meccanica, a dare una certa pendenza alla faccia superiore dei pistoni. Se tale forza è presente, la sua influenza si fa sentire solamente nella fase finale del ciclo di trasferimento. Le forze di origine termica compaiono perché gli elettroni ricevono calore dal reticolo cristallino del silicio e sono cosí liberi di muoversi casualmente. Nel loro moto casuale essi tendono a muoversi verso zone di energia elettronica minima. Questo tipo di forza prevale nella parte finale del ciclo solamente se le forze di campo sono assenti.

Per quanto queste tre forze influiscano sullo spostamento di una frazione relativamente piccola dell'intero pacchetto di carica, esse sono importanti perché nei dispositivi di impiego pratico si può tollerare solo una piccolissima perdita di trasferimento. Per esempio, se a ogni trasferimento si perdesse l'1% di carica, gran parte del pacchetto di carica risulterebbe perduto dopo solo 100 trasferimenti. In generale, la perdita di trasferimento deve essere compresa intorno a una parte su diecimila per essere considerata accettabile nella maggior parte delle applicazioni pratiche. Tuttavia, malgrado questo vincolo, è possibile avere a disposizione dispositivi che possono funzionare a frequenze fino a 100 megahertz se la loro struttura è sufficientemente piccola. Facendo uso delle attuali tecnologie costruttive microelettroniche è possibile progettare e costruire una cella di dispositivo ad accoppiamento di carica con una dimensione laterale fino a un centesimo di millimetro.

Celle unitarie di tali dimensioni sono consentite dalla struttura semplice dei dispositivi ad accoppiamento di carica, che non richiedono di effettuare contatti col silicio nella regione del reticolo. Questo costituisce un'ulteriore differenza con la tecnologia convenzionale dei transistori, che in generale impone numerosi contatti per ogni cella

In questo modello meccanico i pistoni sono di tipo asimmetrico allo scopo di rappresentare il principio di funzionamento di un sistema a due fasi. Ora il fluido si sposta sempre verso destra, indipendentemente dal verso di rotazione dell'albero a gomiti. Rispetto al dispositivo ad accoppiamento di carica il fluido rappresenta un accumulo di elettroni; i pistoni, l'energia potenziale legata alle tensioni applicate; l'albero a gomiti e le bielle, le tensioni pilota, cioè di *clock*, e il loro andamento nel tempo.

Le sezioni trasversali a destra rappresentano elementi ad accoppiamento di carica di un sistema a tre fasi, mentre le curve a sinistra danno l'andamento nel tempo delle relative forme d'onda delle tensioni di *clock*. Il dispositivo consta di un substrato di silicio di tipo *p* (in cui i portatori di segnale sono gli elettroni) che ha sulla superficie uno strato isolante di biossido di silicio. Gli elettrodi conduttori sono posti sulla superficie isolante. All'interno del semiconduttore, sotto lo strato isolante, la conducibilità del silicio può essere modificata per formare uno strato di tipo *n* (in cui i portatori di segnale sono le lacune). Quando si applica agli elettrodi la tensione di *clock*, alcuni elettroni in prossimità di ciascun elettrodo formeranno un pacchetto di carica discreto (*punti neri*) che si sposta verso destra a ogni intero ciclo. In realtà i pacchetti si spostano verso destra per effetto dello spostamento laterale del « pozzo di potenziale » locale in cui vengono a trovarsi (*contorni bianchi nel substrato*).

Le sezioni trasversali a destra mostrano tre elementi ad accoppiamento di carica del tipo a due fasi, e le curve a sinistra danno ancora l'andamento nel tempo delle relative tensioni di *clock*. In questo caso ai pozzi di potenziale viene fornita l'asimmetria richiesta mediante l'introduzione di regioni di tipo *n* a diversa conducibilità immediatamente al di sotto dello strato isolante. Analogamente a quanto mostrato nella figura in alto, le tensioni esterne applicate agli elettrodi provocano il movimento verso destra dei pacchetti di elettrodi attraverso lo strato di tipo *n*. Le curve bianche indicano l'andamento dei potenziali nel substrato.

Sezioni normali alla superficie del semiconduttore che mostrano gli andamenti dell'energia potenziale per un caratteristico elemento di memoria ad accoppiamento di carica all'interno della massa (per meglio rappresentare i pozzi di potenziale, questo diagramma è ruotato di 90° rispetto a quelli della pagina precedente). Nello schema in alto è rappresentata la distribuzione di carica quando nel pozzo di potenziale non vi sono elettroni; nello schema in basso, alcuni elettroni occupano il pozzo. Si può osservare che il pozzo è piú profondo quanto maggiore è la tensione di *clock*.

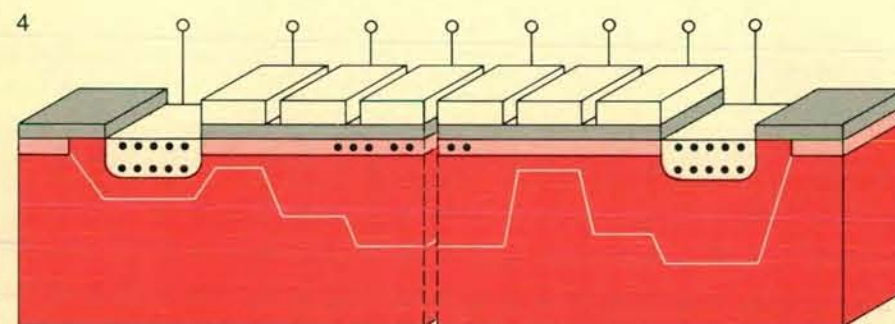
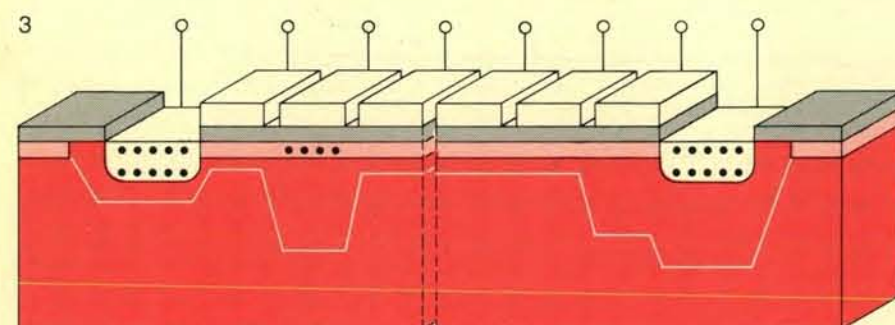
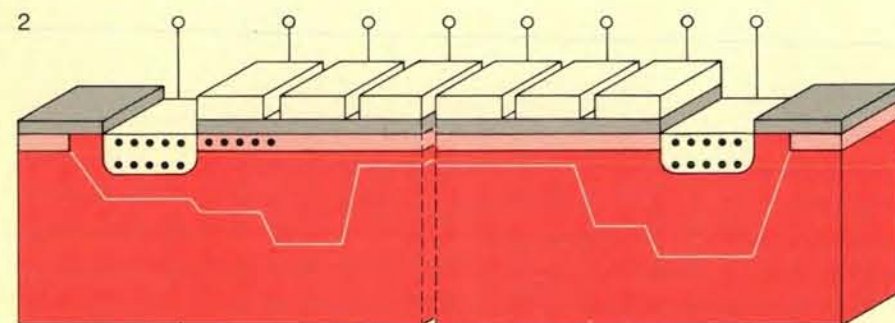
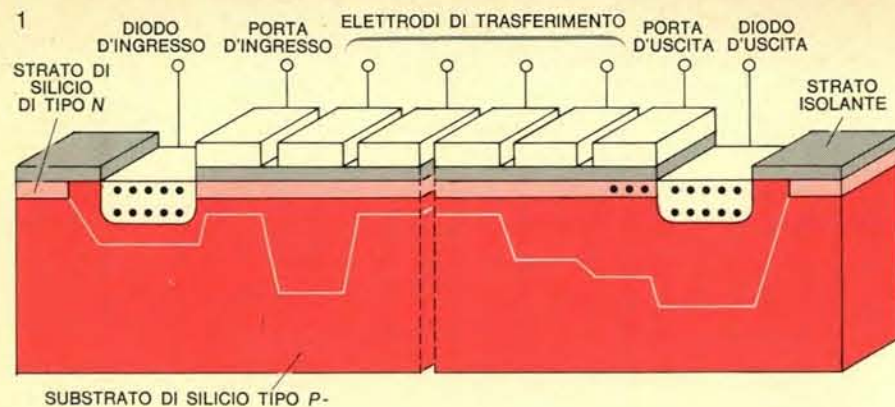
funzionale. Tali contatti consumano una quantità considerevole di silicio necessario sia per l'area di contatto, sia per le tolleranze richieste per realizzare un buon collegamento elettrico. Dal punto di vista costruttivo questo aspetto, piú di ogni altro, rende particolarmente attraenti i nuovi dispositivi ad accoppiamento di carica.

La capacità di creare, spostare e rivelare numerosi pacchetti separati di elettroni in una piccola piastrina di materiale semiconduttore suggerisce che il principio di accoppiamento di carica può essere sfruttato nell'assolvimento di numerosi compiti nel settore del trattamento delle informazioni. In particolare, la caratteristica dei dispositivi ad accoppiamento di carica di manipolare in maniera fortemente ordinata dei pacchetti di carica, favorisce il loro impiego quali sensori di immagine, elementi di memoria dei calcolatori e elaboratori di segnali campionati. In ciascuna applicazione, la funzione viene svolta mediante un'opportuna combinazione di celle unitarie ad accoppiamento di carica che funzionano ognuna esattamente come è stato descritto precedentemente.

Il silicio, che è il materiale con cui sono costruiti piú comunemente i dispositivi ad accoppiamento di carica è molto sensibile alle radiazioni visibili e del vicino infrarosso (*si veda la figura a pagina 23*). Ciò significa che quando la luce colpisce un substrato di silicio, la radiazione viene assorbita (tramite l'effetto fotoelettrico di Einstein) e dà origine alla generazione di elettroni in quantità proporzionale alla quantità di luce incidente. Se nel silicio esiste una distribuzione di « pozzi di potenziale » quale quella dei dispositivi ad accoppiamento di carica, questi elettroni andranno a riempire i pozzi fino a un livello corrispondente alla quantità di luce in loro prossimità. Tale creazione di elettroni per via elettroottica rappresenta un ingresso al dispositivo ad accoppiamento di carica completamente differente da quello relativo al « registro di scorrimento » descritto precedentemente e rende il concetto dell'accoppiamento di carica adatto a diversissimi tipi di applicazioni. Nondimeno, i pacchetti di elettroni generati per via elettroottica possono essere spostati, come nel « registro di scorrimento », fino a un elemento di rivelazione e convertiti in un segnale elettrico rappresentativo dell'immagine ottica incidente sulla superficie. Questo segnale, opportunamente trattato, può essere visualizzato su un tubo a raggi catodici. In tal modo un dispositivo ad accoppiamento di carica può fungere perfettamente da telecamera.

Uno dei piú significativi vantaggi dei sensori di immagine ad accoppiamento di carica rispetto ai sensori a tubo a vuoto consiste nella possibilità di individuare con precisione la localizzazione degli elementi fotosensibili l'uno rispetto all'altro. In una telecamera a tubo a vuoto l'immagine video viene « letta » da un materiale fotosensibile mediante la scansione di un raggio elettronico. La posizione del raggio non può mai essere esattamente nota a causa dell'imprecisione dei circuiti di scansione dovuta al rumore elettrico casuale. In un sensore ad accoppiamento di carica la posizione dei singoli elementi fotosensibili è nota esattamente, poiché è definita al momento della fabbricazione del componente. Tale precisione « metrica » è importante per realizzare il migliore allineamento nelle telecamere a colori e nelle applicazioni che richiedono la riduzione dei dati dell'immagine ottenuta, come nelle missioni spaziali fotografiche e in fotogrammetria.

È solitamente conveniente, ai fini della loro trattazione, separare i sensori ad accoppiamento di carica in due categorie: sensori lineari e sensori superficiali. Un sensore di immagine lineare è una semplice disposizione in linea retta di fotosensori con i relativi circuiti di lettura. Un sensore di immagine superficiale è costituito invece da un mosaico bidimensionale di fotosensori, sempre collegati ai relativi circuiti di lettura.

I sensori di immagine lineari sono utilizzati in numerose applicazioni, incluse le trasmissioni di immagini terra-aria e spazio-aria, la registrazione in facsimile e nella televisione a scansione lenta. L'immagine da visualizzare viene ricavata con uno spostamento relativo tra il sensore e la scena da riprendere mantenendo l'asse della fila di sensori perpendicolare alla direzione dello spostamento. In questi apparati viene comunemente richiesta una risoluzione di 500 o piú elementi fotosensibili. Un dispositivo elementare di questo tipo può constare semplicemente di un « registro di scorrimento » ad accoppiamento di carica e di un diodo di uscita. In tale struttura l'immagine si ottiene mantenendo stazionari i « pozzi di potenziale », arrestando i *clock* di tensione per un certo periodo di tempo (il « tempo di integrazione »)

La sequenza di schemi rappresenta il funzionamento in ingresso e in uscita di un « registro di scorrimento » ad accoppiamento di carica a tre fasi, costituito da otto elementi. Il segnale penetra nel dispositivo attraverso un diodo di ingresso, che funge da sorgente di elettroni. Se la « porta » di ingresso viene mantenuta a bassa tensione, nessun elettrone può percorrere il « canale » di energia potenziale (*1*). Allo scopo di immettere gli elettroni nel dispositivo, si dovrà attendere che gli elettrodi della fase 1 siano in condizioni di alta tensione e quindi « aprire » la « porta » innalzando la sua tensione (*2*). Gli elettroni occuperanno allora il « pozzo di potenziale » fino a che il livello di energia degli elettroni nel pozzo non eguaglierà quello degli elettroni della sorgente. A questo punto la tensione della « porta » di ingresso può venire abbassata per isolare la sorgente (*3*), e il pacchetto di carica creato può essere fatto proseguire lungo il canale (*4*). Il segnale è rivelato trasferendo il pacchetto a un diodo di uscita.

e successivamente estraendo l'informazione facendo ripartire i *clock*. Questo semplice dispositivo può essere praticamente utilizzato solamente in applicazioni speciali che consentano tempi di integrazione molto lunghi. Il motivo di ciò è dovuto alla « sfocatura » dell'immagine che si verifica quando l'impulso di *clock* entra nel « registro di scorrimento » allo stesso tempo in cui esso viene illuminato.

Un sensore di immagine lineare ad accoppiamento di carica di reale utilità pratica è piú complesso. Esso consta di una fila di fotosensori che funge da accumulatore della distribuzione di fotocariche, piú un « registro di scorrimento » avente un elemento ad accoppiamento di carica in corrispondenza di ogni elemento fotosensibile per essere in grado di spostare i pacchetti di carica risultanti fino all'uscita. Gli elementi della fila di fotosensori sono singoli dispositivi ad accumulazione di carica con un elettrodo comune chiamato « fotoporta ». Essi risultano isolati elettricamente fra loro da una regione tipo *p* a elevata concentrazione che viene chiamata arresto del canale. La fila di fotosensori è separata dal « registro di scorrimento » da una regione su cui è un elettrodo chiamato « porta di trasferimento ».

Durante il funzionamento la tensione della « fotoporta » viene mantenuta elevata e la carica generata dalla radiazione incidente (fotocarica) viene trasmessa ai singoli elementi fotosensibili. Al termine del tempo di integrazione la tensione della « porta di trasferimento » viene portata al suo valore normale di bassa tensione. Gli elettrodi del « registro di scorrimento » ad accoppiamento di carica adiacente agli elementi fotosensibili sono pure portati ad alta tensione. La tensione della « fotoporta » viene a questo punto abbassata e la carica fotoelettrica accumulata si trasferisce al « registro di scorrimento ». Quando il trasferimento è completato, la tensione della « porta di trasferimento » viene abbassata e la tensione della « fotoporta » è riportata al suo elevato valore normale per dare inizio a un nuovo periodo di integrazione. Contemporaneamente il « registro di scorrimento » ad accoppiamento di carica viene sottoposto alla tensione di *clock* per ottenere la lettura della distribuzione di carica.

Il progetto piú economico di un sensore di immagine ad alta densità prevede un « registro di scorrimento » a ogni estremità del reticolo fotosensibile. Essendo necessario un elemento ad accoppiamento di carica per ogni elemento fotosensibile, la distanza fra gli elementi fotosensibili è uguale alla

distanza fra gli elettrodi del « registro di scorrimento » in un registro costituito da dispositivi ad accoppiamento di carica a due fasi ed è uguale ad una volta e mezzo la distanza fra gli elettrodi del « registro di scorrimento » in un registro a tre fasi. Nell'esempio visto, il segnale proveniente dai due « registri di scorrimento » a tre fasi è trasportato su un registro a tre fasi e due elementi per alimentare un preamplificatore integrato. Se invece si utilizzano dispositivi a due fasi, è possibile inviare la carica che contiene il segnale direttamente a un diodo di uscita, che costituisce contemporaneamente l'ingresso del preamplificatore integrato. Questi esempi rendono evidente che la struttura ad accoppiamento di carica a due fasi è non solo piú facile da innescare, ma anche di realizzazione piú economica nella pratica. Anche se la sua fabbricazione presenta alcune difficoltà a causa dell'asimmetria richiesta, sarà probabilmente questa la struttura dominante quando i dispositivi ad accoppiamento di carica saranno stati perfezionati.

Con un sensore di tipo lineare è possibile riprodurre immagini convenzionali bidimensionali (*si veda la figura alla pagina seguente*). L'immagine da registrare viene fissata a un tamburo rotante che viene fatto muovere rispetto al dispositivo; la velocità di rotazione è sincronizzata con la scansione verticale dello schermo video. Il sensore di immagine ad accoppiamento di carica di tipo lineare fornisce, a ogni periodo, una riga orizzontale sullo schermo video; la rapida ripetizione del periodo dà origine a tutte le righe orizzontali di un quadro completo. In parecchie applicazioni l'elemento mobile del sistema è rappresentato dal dispositivo sensore, come nella ricognizione aerea, in cui il dispositivo viene fissato all'aereo o al satellite.

La qualità delle immagini ottenute con un sensore lineare ad accoppiamento di carica è eccellente, grazie all'elevato rapporto dinamico risultante (*si veda la figura a pagina 25*). Per rapporto dinamico si intende il rapporto tra la massima e la minima intensità rilevabile dell'immagine. La qualità della riproduzione dimostra che è già possibile, con gli attuali dispositivi ad accoppiamento di carica, raggiungere altissimi rendimenti di trasferimento e ridurre a bassi valori il rumore elettrico.

Confronto tra le caratteristiche di assorbimento spettrale di un elemento fotosensibile ad accoppiamento di carica (*curva in colore*) e dell'occhio umano (*curva in nero*). Il materiale semiconduttore è sensibile non solo alla luce visibile (da 0,4 a 0,7 micron), ma anche alla radiazione del vicino infrarosso (da 0,7 a 1,1 micron). L'assorbimento di tale radiazione dà origine, nel substrato di silicio, alla generazione di elettroni in quantità proporzionale all'intensità della radiazione incidente. Questa proprietà elettroottica permette l'impiego dei dispositivi ad accoppiamento di carica come sensori di immagine.

I sensori di immagine superficiali trovano il loro principale campo di impiego come camere da ripresa televisiva. L'immagine viene ricavata dalla scansione convenzionale, riga per riga, del mosaico reticolare di sensori e dalla successiva riproduzione del segnale video sullo schermo di un tubo a raggi catodici mediante scansione standard. Un sensore superficiale ad accoppiamento di carica che dovrà operare in questo modo, può essere progettato secondo uno schema analogo a quello del sensore lineare. Come nel normale processo di trasmissione televisiva, l'immagine viene letta dividendola in due campi separati, di cui il primo è costituito da tutti gli elementi fotosensibili con numero pari di ogni colonna, e il secondo da quelli dispari, piuttosto che leggere gli elementi dispari e pari in parallelo, come accade nei sensori di immagine lineari.

Il sensore di immagine superficiale funziona come segue. La radiazione luminosa che colpisce ininterrottamente le aree fotosensibili libera elettroni che si accumulano in pacchetti di carica nei « pozzi di potenziale » creati dalla tensione della « fotoporta ». Dopo un trentesimo di secondo i pacchetti di carica raccolti nei fotosensori adiacenti agli elettrodi della fase 1, vengono trasferiti nella regione sottostante agli elettrodi della fase 1 che hanno elevato il loro potenziale mentre è sceso il potenziale delle « fotoporte ». I pacchetti di carica raccolti nelle aree fotosensibili adiacenti agli elettrodi della fase 2, invece, non subiscono il trasferimento poiché la tensione della fase 2 rimane bassa. Al termine del trasferimento alla fase 1, la tensione delle « fotoporte » ritorna al suo normale valore elevato cosicché altri elettrodi ricominciano ad accumularsi nelle aree fotosensibili esaurite. I pacchetti di carica contenuti nel « registro di scorrimento opacizzato » vengono ora trasferiti al « registro di scorrimento orizzontale » in cima al reticolo. Ogni trasferimento verticale riempie il registro orizzontale che trasmette infine all'uscita l'intera informazione rappresentante una riga dell'immagine video. Quando tutte queste righe sono « lette » in questo modo (l'intera procedura richiede solo un sessantesimo di secondo) entrano in azione i fotosensori adiacenti agli elettrodi della fase 2 trasmettendo in maniera analoga all'uscita l'informazione relativa al secondo campo di immagine. In conclusione, l'intera operazione inizia e si completa nel tempo di un trentesimo di secondo.

Un sensore di immagine sperimentale costruito per funzionare nel modo descritto consta di una griglia rettangolare formata da 100 per 100 elementi fotosensibili (*si veda la figura a pagina 17*). Ognuno di essi, col relativo elettrodo rivelatore, occupa un'area di un millesimo di millimetro quadrato. Tutti i 10 000 elementi stanno su una piastrina di 3 × 4 millimetri. Un'immagine ripresa con un simile apparato può essere rappresentata su uno schermo televisivo, come se provenisse da una normale telecamera.

Questo e altri simili sensori di immagine ad accoppiamento di carica sono ancora in una fase sperimentale, ma indicano già una promettente via per giungere alla realizzazione di telecamere efficaci. L'affidabilità della tecnologia a stato solido, il funzionamento a bassa tensione, la limitata dissipazione di potenza, l'elevato rapporto dinamico, la esatta riproducibilità metrica e la risposta alla radiazione visibile e del vicino infrarosso rappresentano per il potenziale utilizzatore sostanziali vantaggi rispetto ai sensori di immagine a tubo a vuoto e a stato solido.

Il principio dell'accoppiamento di carica appartiene fondamentalmente all'elettronica dei semiconduttori piuttosto che all'elettroottica: tuttavia, grazie alle caratteristiche elettroottiche del silicio, le proprietà fotosensibili dei dispositivi ad accoppiamento di carica hanno inizialmente prevalso nelle applicazioni tecnologiche. Ma non per questo devono essere trascurate le proprietà di tali dispositivi nel trattamento delle informazioni.

Una fila di semiconduttori ad accoppiamento di carica costituisce un « registro di scorrimento analogico » a campionamento temporale, teoricamente ideale. Ciò significa, in elettronica, la realizzazione di una linea di ritardo in cui il ritardo è proporzionale al rapporto tra velocità di accettazione e velocità di lettura dell'informazione; se la fila è abbastanza lunga da contenere l'intero messaggio, le velocità di accettazione e rivelazione possono differire e il massimo ritardo ottenibile è limitato solamente dalla generazione di elettroni casuali di origine termica. A basse temperature si possono raggiungere parecchi minuti di ritardo.

Le file di dispositivi ad accoppiamento di carica possono essere impiegate come elementi di immagazzinamento di dati numerici ad accesso sequenziale o come componenti ibridi del tipo a tamburo o a disco. Un notevole vantaggio nel loro impiego come elementi di memoria consiste nell'eliminazione di ogni sorta di movimento meccanico.

La natura analogica intrinseca ai pacchetti di carica di un dispositivo ad accoppiamento di carica rende possibile un vasto settore di applicazioni potenziali come elaboratori di segnali campionati. In linea teorica l'impiego di un dispositivo ad accoppiamento di carica come sensore di immagini non è altro che una particolare applicazione del dispositivo come « registro di scorrimento analogico ». Se si restringe il campo dei dispositivi elaboratori di segnali campionati a quelli con ingresso puramente elettrico, escludendo quindi quelli a ingresso ottico, si trovano come membri principali le linee di ritardo variabile e i filtri.

Una linea di ritardo è un circuito che riproduce il piú esattamente possibile un segnale di ingresso ritardandolo di una quantità di tempo finita. Una linea di ritardo si dice « variabile » se il tempo di ritardo può essere variato elettricamente. Il dispositivo ad accoppiamento di carica agisce naturalmente come linea di ritardo poiché qualsiasi segnale applicato al suo diodo di ingresso comparirà all'uscita (in forma campionata) dopo l'intervallo di tempo richiesto dai pacchetti di carica per scorrere lungo tutti gli elementi della struttura. Il dispositivo ad accoppiamento di carica può essere impiegato come linea di ritardo in parecchi modi: nel primo, il piú semplice, il ritardo è uguale al numero di celle unitarie divise per la frequenza di *clock* del dispositivo; diversamente, quando i dati compaiono come scariche di impulsi, il registro ad accoppiamento di carica può essere caricato durante la emissione della scarica e i dati conservati per il tempo voluto e infine rivelati. In questo modo il dispositivo può compiere le funzioni di una memoria di accumulo (*buffer*).

Una linea di ritardo ad accoppiamento di carica presenta numerosi vantaggi rispetto alle linee convenzionali su vetro e anche rispetto a piú raffinati dispositivi a onda acustica superficiale (si veda l'articolo *Le onde acustiche di superficie* di Gordon S. Kino e John Shaw, in « Le Scienze », n. 53, gennaio 1973). Tra questi stanno l'ampio rapporto dinamico (oltre 60 decibel dopo 30 millisecondi a temperatura ambiente) e la possibilità di regolazione elettronica separata della velocità di propagazione e del tempo di ritardo. Linee di ritardo con questi requisiti saranno di grande utilità nel settore delle comunicazioni e delle applicazioni della televisione e renderanno piú facili gli attuali metodi di produzione di ritardi di segnale controllati. Un'applicazione di particolare interesse riguarda il «convertitore del rapporto di scansione» spesso utile nelle comunicazioni video. Il dispositivo a carica accoppiata funziona in questo caso come *buffer* con le caratteristiche esaminate precedentemente e può quindi accettare e in un secondo tempo rivelare le immagini video con diverse frequenze di scansione in modo da rendere possibile l'accoppiamento fra le ampiezze di banda effettive del sistema di trasmissione e le necessità dei diversi standard della riproduzione televisiva.

Questo schema rappresenta uno dei possibili apparati per la riproduzione di immagini bidimensionali con dispositivi sensori lineari ad accoppiamento di carica. L'immagine da riprodurre viene fissata a un tamburo rotante (*a sinistra*) la cui velocità di rotazione è sincronizzata con quella di scansione verticale di uno schermo televisivo (*a destra*). Il dispositivo ad accoppiamento di carica e il suo circuito di lettura generano le righe orizzontali a una velocità sufficiente a riprodurre sullo schermo l'immagine.

Queste fotografie danno un'idea dell'elevato livello di fedeltà ottenibile con un sensore di immagine costituito da 500 elementi ad accoppiamento di carica in diverse condizioni di illuminazione. A sinistra è la fotografia originale; la fotografia al centro rappresenta lo schermo televisivo alimentato dal sensore ad accoppiamento di carica in condizioni perfette; la fotografia a destra rappresenta l'immagine sullo schermo con illuminazione ridotta di mille volte; per produrre questa immagine, il dispositivo ad accoppiamento di carica dovette spostare pacchetti di circa 400 elettroni attraverso un centimetro di silicio senza dispersione.

Partendo dal semplice principio di linea di ritardo è possibile giungere a dispositivi piú complessi di elaborazione di segnale. Se, infatti, in alcuni punti di una linea di ritardo vengono introdotte prese in cui il segnale viene misurato e rinviato a uno o piú stadi precedenti in modo da influire sulla trasmissione dei dati, si è realizzato un dispositivo utilizzabile come filtro. Una simile struttura può essere paragonata a un filtro passabanda nel quale la frequenza di risonanza del circuito è funzione diretta della frequenza di *clock*. Con un simile dispositivo è stato raggiunto un miglioramento nel rapporto segnale/rumore nell'ambito di un decibel del massimo valore teorico.

Filtri accordati trovano applicazione nelle comunicazioni a largo spettro e nei radar per rivelare segnali deboli in ambienti ad elevato rumore. In queste applicazioni i dispositivi ad accoppiamento di carica potranno essere integrati ai dispositivi a onde acustiche superficiali, che possono normalmente essere impiegati solo per ritardi inferiori ai 100 microsecondi.

Come è stato già detto, un elemento di memoria ad accoppiamento di carica può immagazzinare pacchetti di elettroni aventi carica diversa, a seconda delle caratteristiche di progetto e di funzionamento della cella unitaria. Ciò nonostante, è possibile assegnare convenzionalmente un valore alla capacità di accumulazione di carica della cella e trattare il dispositivo come un componente numerico binario. Per esempio, si potrebbe stabilire che se un elemento contiene una quantità di carica inferiore alla metà della sua carica di saturazione, esso si trova al livello « zero », mentre se possiede una carica maggiore della metà della carica di saturazione, esso si trova a livello « uno ». In tal modo, l'elemento di accumulazione diventa un « bit » di memoria e una linea di ritardo ad accoppiamento di carica può essere fatta funzionare come « registro di scorrimento numerico » o come memoria ad accesso seriale. Poiché questa stessa funzione può essere svolta da altri dispositivi, ci si potrebbe chiedere quali siano i vantaggi dei dispositivi ad accoppiamento di carica rispetto agli altri. La risposta sta nel rapporto costo/efficacia. Una memoria ad accoppiamento di carica possiede non solo gli stessi pregi di un componente tradizionale a semiconduttori (compatibilità con altri elementi elettronici, nessun elemento in movimento, bassa potenza e tensione, velocità variabili di *clock* e altre simili caratteristiche), ma offre in piú un basso costo per bit confrontabile a quello di una memoria magnetica. Ciò dipende dalla particolare semplicità strutturale del dispositivo ad accoppiamento di carica: questo permette di introdurre reticoli di memoria di circa 250 000 bit per componente su una piastrina di silicio di dimensioni inferiori a un centimetro.

In aggiunta, la potenza assorbita da un dispositivo di memoria ad accoppiamento di carica è molto bassa in quanto l'elemento di memoria non è attivo. La potenza necessaria per spostare la carica accumulata su un elemento ad accoppiamento di carica in un altro elemento adiacente in un microsecondo è di circa un microwatt. Inoltre, non sempre in una memoria occorre spostare contemporaneamente tutti i bit. Pertanto, un dispositivo di memoria contenente un milione di bit alla frequenza di un milione di hertz assorbirebbe una potenza compresa fra un milliwatt e 1 watt, in un volume inferiore a quello di un pacchetto di sigarette.

Un ulteriore pregio risiede nel fatto che il dispositivo ad accoppiamento di carica è di tipo fondamentalmente analogico. Ciò lo rende adatto ad accumulare piú di un bit per ogni cella di memoria realizzando un certo numero di livelli discreti di carica in ogni cella, e incrementando cosí notevolmente la densità di informazione. Per esempio, un dispositivo a 100 000 celle in cui siano stati realizzati 8 livelli di carica è paragonabile a una memoria convenzionale di 300 000 bit. Una simile memoria sarebbe di grande utilità nei convertitori analogico-numerici.

L'insieme di queste caratteristiche fanno ragionevolmente ritenere che le memorie a dispositivi ad accoppiamento di carica daranno inizio a una rivoluzione nella tecnologia dei grandi calcolatori che li renderà accessibili a sempre piú larghi strati di utilizzatori nella vita di ogni giorno. Il principio di accoppiamento di carica costituisce, in breve, una innovazione fondamentale nell'elettronica dei semiconduttori. Per la sua semplicità di progetto e di costruzione, le sue elevate prestazioni in termini di rapporto dinamico e basso assorbimento di potenza, e per la sua elevata densità di impaccamento a basso costo, la tecnologia dell'accoppiamento di carica creerà nuove e importanti applicazioni per i semiconduttori, con un impatto diretto nella nostra vita.

In questa fotografia al microscopio elettronico, ingrandita 250 000 volte, si possono vedere alcune particelle a forma di proiettile del virus della stomatite vescicolare (VSV), che provoca una lieve malattia nell'uomo, ma una grave malattia in alcuni animali. La particella rotondeggiante a sinistra rispetto al centro della fotografia è un virus a interferenza difettosa. Esso, nel suo genoma, ha solamente una parte di acido ribonucleico (RNA) ed è, quindi, dipendente dal virus completo per la sua riproduzione. I cordoni attorcigliati tra le particelle virali sono nucleoproteine che si sono staccate dalle particelle virali durante la preparazione per l'esecuzione della fotografia. Queste spire di nucleoproteine contengono l'acido ribonucleico contenente il genoma del virus. Le microfotografie sono state riprese da Ronald Leavitt dell'Università della California a San Diego.

Virus lenti, asintomatici e ricorrenti

Alcuni virus possono rimanere nel loro ospite senza che si manifestino sintomi di infezione. Si ritiene che tali virus possano causare o innescare malattie croniche degenerative nell'uomo

di John J. Holland

Negli ultimi anni è stato chiarito che alcune malattie ad andamento lento e continuo, che non sembrano apparentemente essere di origine infettiva, sono causate o iniziate da virus ad azione insolitamente lenta. È stato dimostrato che quattro gravi malattie neurologiche sono provocate da virus lenti: il kuru e la malattia di Creutzfeldt-Jakob nell'uomo, lo *scrapie* della pecora e l'encefalopatia contagiosa nel visone. Vi sono molte prove preliminari, finora non conclusive, che nell'uomo un certo numero di comuni malattie degenerative, come per esempio la sclerosi multipla, l'artrite reumatoide, la leucemia e il diabete, possano in realtà essere la conseguenza di infezioni virali lente e non manifeste. Se le prove preliminari del ruolo dei virus nella eziologia delle malattie degenerative venissero ulteriormente confermate, praticamente tutte le malattie croniche e lentamente progressive dell'uomo, che sono tuttora di origine sconosciuta, dovrebbero essere analizzate sistematicamente per determinare la loro possibile origine virale.

In questo tipo di malattie da virus la lentezza non è necessariamente una caratteristica del virus per se stesso, ma piuttosto del decorso della malattia. In una tipica infezione acuta virale, la cellula infettata generalmente muore dopo 10-48 ore dall'infezione. Nell'infezione da virus lenti, il virus può vivere e riprodursi nelle cellule senza provocare la morte della cellula ospite. La presenza del virus viene però avvertita dal sistema immunitario dell'organismo ospitante la cui reazione all'infezione virale può essere responsabile di buona parte delle alterazioni riscontrate nelle malattie degenerative.

In un'infezione virale acuta, quale l'influenza o la poliomielite, il virus invasore soverchia l'attività cellulare e inizia a sintetizzare nuovi componenti virali. Alla fine un ammasso di nuovi virus esplode fuori dalla cellula, che a questo punto muore, mentre le nuove particelle virali continuano a infettare nuove cellule. Tuttavia non tutte le infezioni virali terminano con la morte della cellula ospite. Da molto tempo si sa che il virus dell'*Herpes simplex*, che provoca ragadi e bolle infiltrate, può rimanere inattivo e asintomatico nell'ospite, per mesi, anni o persino decenni. A Los Angeles, Jack G. Stevens della Facoltà di medicina dell'Università della California, ha dimostrato che il virus dell'*Herpes* rimane latente entro le cellule dei gangli nervosi senza produrre alcun sintomo, finché un fattore scatenante (come per esempio una scottatura solare, una febbre o una esposizione al freddo intenso) rende attivo il virus e si manifestano piccole eruzioni cutanee. Tuttavia, tra una esacerbazione e l'altra apparentemente il virus non distrugge le cellule nervose entro le quali continua a vivere.

Come possono alcuni virus vivere nelle cellule e moltiplicarsi senza provocare la morte della cellula ospite? Parte della risposta risiede nel fatto che tali virus non inibiscono la sintesi dell'acido ribonucleico (RNA) e delle proteine, e questo fa sí che la cellula infettata continui a funzionare. D'altra parte, i virus lenti manifestano quasi sempre la capacità di alterare la composizione delle proteine che fanno parte della superficie della membrana esterna della cellula infettata. Sebbene molti virus responsabili di malattie acute siano capaci di fare la stessa cosa, dal punto di vista dell'ospite una superficie cellulare alterata in una cellula che sopravvive è molto piú pericolosa che in una cellula morente. Quando una cellula alterata resta viva e si moltiplica, essa e le cellule discendenti hanno forti probabilità di essere attaccate quali oggetti estranei dal sistema immunitario. Il risultato può essere una malattia: neurologica, auto-immunitaria o un cancro.

I mixovirus (i virus dell'influenza) e i paramixovirus (i virus del morbillo, della parotite e di malattie di questo gruppo) hanno un complesso di geni che rappresentano i codici necessari per la formazione di diverse proteine virali. Alcune di queste proteine hanno una forte affinità per la membrana esterna della cellula infettata. Altri geni virali dirigono la sintesi dell'RNA virale. L'RNA di nuova sintesi è ordinatamente sistemato lungo la proteina codificata dal virus, sotto forma di piccoli tubi cilindrici all'interno del citoplasma cellulare. Questi tubuli di ribonucleoproteina (RNP) apparentemente forniscono una intelaiatura protettiva per l'RNA virale. Quando la concentrazione dei tubuli di RNP aumenta, finisce coll'invadere la maggior parte del citoplasma cellulare. Alcuni tubuli di RNP si spostano verso aree della membrana cellulare che contengono proteine virali. Quando queste proteine della membrana, codificate dai geni virali incominciano a staccarsi dalla cellula, l'RNP viene incorporato dalla parte interna della porzione che si sta staccando. Questa racchiude l'RNP portatore del genoma virale, complesso dei geni del virus, quindi, quando esso si stacca, è in grado di infettare un'altra cellula (*si veda l'illustrazione nella pagina seguente*).

Ci si potrebbe aspettare che, poiché i nuovi virus prendono origine dalla membrana cellulare, la parte piú esterna di essi contenga sia proteine cellulari che virali. Utilizzando con il collega E.D. Kiehn una tecnica di marcatura con sostanze radioattive, abbiamo trovato che poco o niente della proteina della cellula ospite viene incorporata nel-

L'infezione di una cellula da parte di un virus a RNA può provocare la formazione di una nuova particella virale per « gemmazione » senza che venga uccisa la cellula ospite. Una particella virale infettiva si attacca a una cellula (*1*) e svuota il suo genoma (RNA situato in una capsula o rivestito di nucleoproteina) nel citoplasma di una cellula (*2*). Una transcriptasi, un enzima presente nel genoma virale, dà origine a numerosi tipi di molecole di RNA, portatrici di messaggio genetico (mRNA), che si attaccano ai ribosomi nel citoplasma. Il mRNA produce una proteina che funziona da involucro, una proteina appuntita, una nucleoproteina a forma di capsula e gli enzimi replicasi e transcriptasi (*3*). La replicasi si combina con l'RNA virale originale per produrre nuovo RNA virale, che viene incorporato in una capsula di nucleoproteine ed emigra verso la membrana cellulare esterna. Nel frattempo anche le proteine involucro e quelle appuntite si sono spostate verso la membrana esterna, dove si formano delle piccole placche (*4*). Il genoma virale formato da RNA e proteine si dispone sotto alle placche più grandi. Queste formano alcune gemme che racchiudono il genoma virale (*5*). Le particelle che si distaccano sono virus completi che possono infettare altre cellule.

la particella di virus influenzale, mentre vengono inglobati i grassi e i carboidrati della cellula ospite. Ciò suggerisce che le proteine formate sulla base del codice genetico virale alterino completamente alcune zone della membrana cellulare. Non è quindi sorprendente che durante le infezioni virali acute, come l'influenza e il morbillo, le cellule cessino di funzionare quando il loro citoplasma si riempie di tubuli di RNP virale e le loro membrane esterne vengono infiltrate da proteine virali.

In una serie di ricerche di avanguardia, iniziate piú di dieci anni fa, Robert Rustigian della Facoltà di medicina della Tufts University, ha dimostrato che l'infezione normalmente letale di cellule allevate in una coltura di tessuto per mezzo del virus del morbillo può diventare non letale in determinate condizioni. Egli riuscí a coltivare cellule permanentemente infettate da questo virus in presenza dei rispettivi anticorpi. Le cellule continuarono a dividersi per centinaia di generazioni e a trasmettere continuamente durante la divisione cellulare il virus del morbillo alle cellule figlie. Sebbene il citoplasma delle cellule contenesse tipiche inclusioni virali e grandi quantità di proteine virali, fu possibile isolare dalle colture solamente piccole quantità di virus infettivi. Rustigian concluse che era stato prodotto nelle cellule un virus incompleto.

Il significato clinico di questa scoperta divenne chiaro nel 1969, quando il virus del morbillo venne isolato nelle cellule del cervello di soggetti affetti da una malattia infiammatoria chiamata panencefalite subacuta sclerotizzante (PESS). Questa rara e fatale malattia conduce a una perdita progressiva delle funzioni mentali e motorie. Ormai da decenni è noto che la PESS generalmente si presenta mesi o anni dopo un caso grave di morbillo e che i pazienti affetti da PESS tendono ad avere un tasso molto elevato di anticorpi per il morbillo. I tentativi di isolare il virus del morbillo dal tessuto cerebrale di pazienti affetti da PESS sono stati senza esito finché queste cellule non sono state allevate assieme a colture di cellule sensibili al morbillo. È chiaro che il genoma del virus del morbillo si trova in forma latente nelle cellule cerebrali dei soggetti affetti da PESS e che esso viene attivato dalla fusione con cellule sensibili. Luiz Horta-Barbosa, John L. Sever, David A. Fuccillo e i loro colleghi presso l'Istituto nazionale per le malattie neurologiche e l'apoplessia cerebrale hanno trovato che virus latenti del morbillo sono diffusi anche nelle cellule dei linfonodi dei pazienti affetti da PESS. Questa scoperta fa pensare che la panencefalite sia una infezione virale generalizzata e di lunga durata, che colpisce una parte dell'organismo piú estesa che non le sole cellule cerebrali.

Non esiste una spiegazione al fatto che la maggior parte delle persone non risenta alcuna conseguenza del morbillo, mentre pochi, rari individui presentino come complicanza una malattia cerebrale progressiva e fatale. È chiaro che i pazienti affetti da questo tipo di encefalite non hanno scarsa capacità di produrre anticorpi, poiché essi hanno una elevata concentrazione di anticorpi per il morbillo. Tuttavia anticorpi capaci di neutralizzare virus adulti al di fuori delle cellule possono essere inefficaci nel combattere la diffusione del genoma virale da una cellula all'altra attraverso ponti citoplasmatici o per mezzo della divisione cellulare. Alcuni ricercatori sospettano che i pazienti affetti da PESS abbiano un difetto immunologico piú sottile: un'incapacità a distruggere le cellule infettate la cui membrana sia alterata dal virus. Naturalmente il virus isolato da pazienti affetti da PESS per mezzo della commistione con cellule sensibili non è veramente rappresentativo dello stato dei virus nelle cellule dei pazienti. Il virus si trova nei pazienti molto probabilmente sotto forma di particelle incomplete di RNP che non sono in grado di infettare. Il virus incompleto probabilmente si riproduce nello stesso modo sbilanciato che Rustigian ha osservato nelle sue colture di cellule infettate con virus del morbillo.

Anche i pazienti affetti da sclerosi multipla tendono ad avere, nel complesso, un tasso di anticorpi per il morbillo piú elevato dei pazienti affetti da altre malattie neurologiche. Nella sclerosi multipla si verifica un progressivo indurimento di placche del cervello e del midollo spinale che provoca una perdita nella coordinazione muscolare. John M. Adams e David T. Imagawa dell'Università della California presso la Facoltà di medicina di Los Angeles hanno osservato formazioni inclusive cellulari del tipo di quelle provocate dal morbillo nelle lesioni cerebrali di pazienti affetti da sclerosi multipla. Tuttavia il virus del morbillo non è stato isolato né con la tecnica della fusione con altre cellule, né in altro modo da pazienti affetti da sclerosi multipla. Se il virus del morbillo è la causa della sclerosi multipla, il virus si deve riprodurre in modo ancora piú imperfetto di quello che succede nella panencefalite sclerotizzante.

Recentemente V. der Meulen dell'Università di Göttingen e Hilary Koprowski e colleghi dell'Istituto di anatomia e biologia Wistar, con l'impiego della tecnica della fusione cellulare, hanno isolato il virus tipo I della parainfluenza dalle cellule del cervello di pazienti affetti da sclerosi multipla. Il virus della parainfluenza è del tutto simile al virus del morbillo, appartenendo entrambi al gruppo dei paramixovirus.

Un'ulteriore prova della partecipazione dei paramixovirus nell'eziologia della sclerosi multipla è la scoperta di tubuli fittamente ammassati in microfotografie di cellule cerebrali provenienti dai pazienti affetti da questa malattia. Il materiale di tipo virale è stato ritrovato solamente nelle piccole lesioni iniziali, mentre non era presente nelle lesioni piú grandi. Questa scoperta è stata fatta da John Prineas del Dipartimento di medicina dell'Università di Sydney; egli avanza l'ipotesi che le lesioni cerebrali della sclerosi multipla possano essere iniziate dal periodico disseminamento da parte di linfociti portatori del virus nei punti piú recettivi della sostanza bianca cerebrale.

Nel 1970 Koprowski e i suoi collaboratori riferirono di aver trovato il virus del morbillo e un piccolo virus a DNA nelle cellule cerebrali di alcuni pazienti affetti da panencefalite. Il virus a DNA era simile a virus noti per causare tumori: il virus del polioma (un virus dei roditori) e il virus SV-40 (un virus delle scimmie). Billie L. Pad-

In queste microfotografie al microscopio elettronico si può osservare un virus che si stacca dalla membrana cellulare. La microfotografia a sinistra, eseguita da Garth Nicholson dell'Istituto Salk per gli studi biologici, mostra una nuova particella di virus dell'influenza che si sta formando per gemmazione. Le aree dense nella membrana cellulare sono placche di proteine virali. La microfotografia a destra, fornita da George J. Todaro e dai suoi collaboratori presso l'Istituto nazionale del cancro, mostra una particella di virus di tipo C che si sta staccando dalla membrana di una cellula di criceto allevata in coltura.

gett, Duard L. Walker, Gabriele M. Zu Rhein e i loro collaboratori presso la Facoltà di medicina dell'Università del Wisconsin avevano già osservato che ci sono un gran numero di formazioni virali simili al virus SV-40 nelle cellule cerebrali di pazienti affetti da un'altra malattia cerebrale lentamente progressiva e fatale: la leucoencefalopatia progressiva multifocale (LPM). Questa è una malattia della sostanza bianca del cervello; i suoi sintomi comprendono sonnolenza, visione offuscata e andatura incerta. La maggior parte delle vittime della LPM hanno una difettosa risposta immunitaria. È significativo che molti di essi siano stati sottoposti in precedenza a un trattamento farmacologico allo scopo di sopprimere la loro risposta immunitaria, sia per prevenire il rigetto di tessuti trapiantati, sia per altri scopi. Tra il 1971 e il 1972 virus del tipo SV-40 sono stati isolati dal tessuto cerebrale di vittime della leucoencefalopatia da parte dei ricercatori del Wisconsin. Questa è la prima prova che il virus SV-40 può essere interessato in questa malattia dell'uomo.

Vi sono altre malattie del sistema nervoso centrale, ad andamento cronico o lentamente progressivo che hanno probabilmente un'origine virale. Forse le più conosciute di queste malattie sono il kuru e la malattia di Creutzfeldt-Jakob. Entrambe sono forme di demenza presenile lentamente progressiva: esse provocano lo sviluppo prematuro dei deficit mentali che si accompagnano alla vecchiaia. In entrambe le malattie vi sono numerosi vacuoli nelle cellule cerebrali e una proliferazione delle cellule gliari che sono le «assistenti» delle cellule nervose. La materia grigia del cervello degenera e diventa di aspetto spugnoso.

Il kuru è stato riscontrato solo tra popolazioni primitive della Nuova Guinea Orientale. Questi indigeni una volta praticavano un cannibalismo rituale che prescriveva di cibarsi del cervello e delle viscere dei loro parenti defunti. Quando il cannibalismo venne messo fuori legge e la pratica cessò, la frequenza del kuru cadde. La malattia di Creutzfeldt-Jakob è pure rara ma è distribuita in vaste zone. D. Carleton Gajdusek, Clarence J. Gibbs jr. e i loro collaboratori presso l'Istituto nazionale per le malattie neurologiche sono riusciti a trasmettere il kuru e la malattia di Creutzfeldt-Jakob a scimpanzé, partendo da tessuto cerebrale omogeneizzato di pazienti che erano morti di queste malattie. Per gli scimpanzé il periodo di incubazione, durante il quale non vi era alcun sintomo di infezione o di malattia, era compreso tra 10 e 24 mesi. La durata della malattia, (che era invariabilmente letale) era compresa tra 1 e 15 mesi. Le scimmie all'inizio sembravano essere resistenti alla malattia ma con osservazioni più prolungate si scoprì che molte specie di scimmie erano sensibili. In una scimmia reso il kuru comparve 8 anni e mezzo dopo l'inoculazione. Gajdusek e Gibbs sono inoltre riusciti a far passare la malattia da uno scimpanzé all'altro. Anche se la trasmissione di queste malattie fornisce una valida prova che esse sono provocate da virus, i virus stessi non sono stati ancora isolati o allevati in colture cellulari.

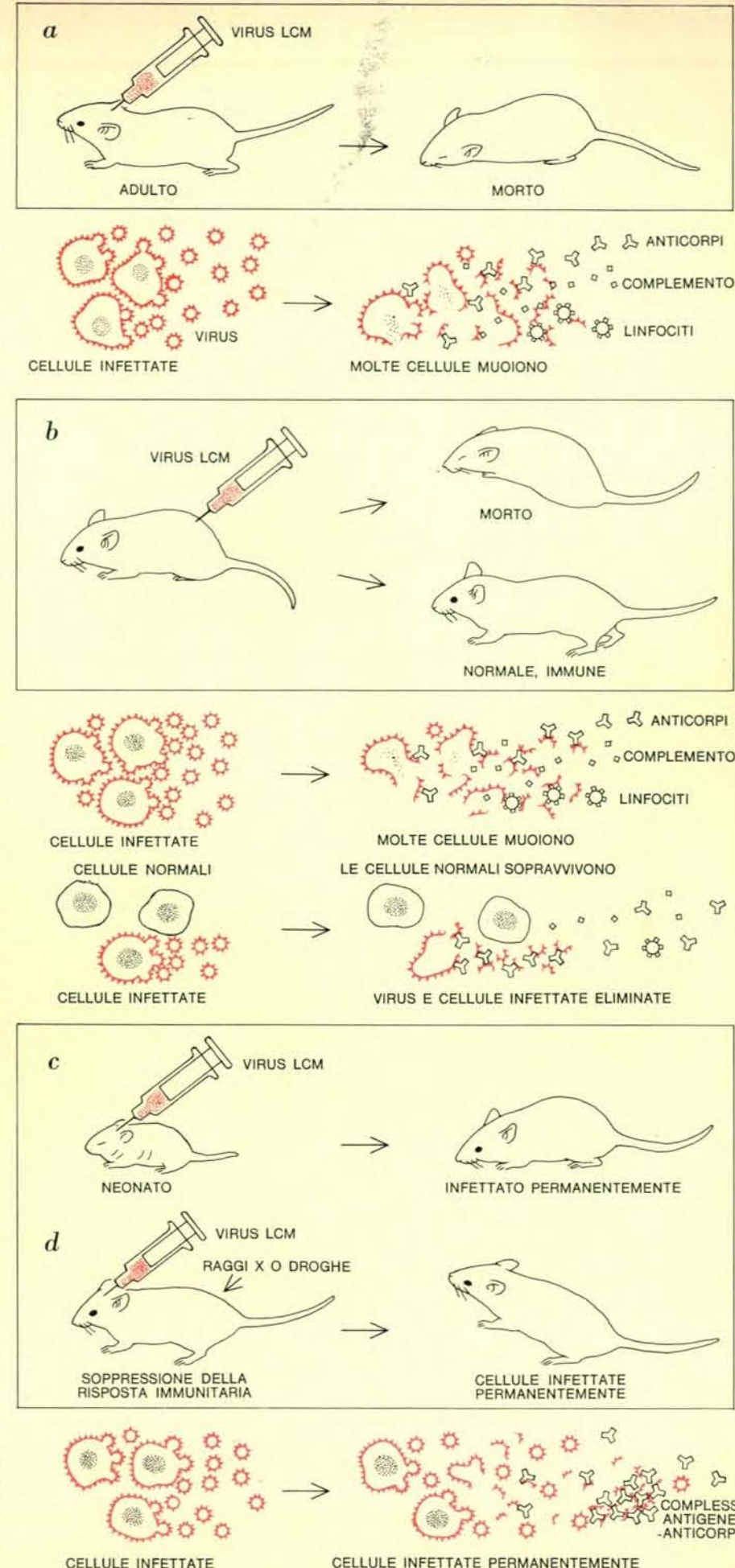
Lo *scrapie* della pecora e l'encefalopatia del visone sono malattie da virus lenti simili al kuru e alla malattia di Creutzfeldt-Jakob. In tutte e quattro queste malattie il virus trasmissibile mostra una sorprendente resistenza al calore e all'inattivazione con radiazioni o con sostanze chimiche. Anche in questo caso nessuno è riuscito a isolare gli agenti dello *scrapie* o dell'encefalopatia del visone, sebbene siano stati fatti molti sforzi.

A causa delle inusuali proprietà degli agenti che trasmettono tutte queste malattie, molti ricercatori ritengono che i fattori causali possano essere non dei virus ma qualche nuovo tipo di agente infettivo. D'altra parte, tutto ciò che si sa delle quattro malattie è in perfetto accordo con la spiegazione che esse siano provocate da tipici virus che sono straordinariamente resistenti alle difese immunitarie dell'organismo. Anche lo sfuggente agente dello *scrapie* è stato probabilmente ormai individuato. H.K. Narang e i suoi collaboratori presso l'Ospedale generale di Newcastle-upon-Tyne, in Inghilterra, recentemente hanno riferito di aver trovato in microfotografie ottenute con il microscopio elettronico densi ammassi di particelle di tipo virale nelle cellule di ratti infettati con lo *scrapie*.

Alcune delle caratteristiche fondamentali delle infezioni da virus lenti sono ben illustrate da una malattia che si riscontra nel topo selvatico in certe zone del mondo e che qualche volta colpisce anche l'uomo. La malattia è la coriomeningite linfocitaria (CML) che provoca un'infiammazione delle membrane che circondano il cervello. Alcune importanti ricerche iniziali su questo virus sono state portate a termine da E. Traub in Germania, il quale ha trovato che, nel topo infettato, si può sviluppare una malattia letale o una forma meno grave seguita da una forte immunità e dalla scomparsa del virus. Tuttavia in pochi topi l'infezione virale continua pur senza che si manifestino segni chiari di infezione.

John E. Hotchin del Dipartimento per la salute dello Stato di New York ha in seguito dimostrato che lo sviluppo dell'infezione cronica del virus della CML dipende da una inibita risposta immunitaria nei confronti del virus. Topi neonati inoculati con il virus sopravvivono perché sono privi di una completa risposta immunitaria. Quando il virus viene iniettato nel cervello di un topo adulto, questo invariabilmente muore; mentre quando viene iniettato in vena il topo può o morire o guarire acquistando un'immunità completa; ciò dipende dal numero di cellule cerebrali che sono infettate dal virus e in seguito uccise dalla risposta immunitaria. Hotchin ha anche dimostrato che, nel topo adulto inoculato con il virus della CML, il trattamento con i raggi X o con i farmaci per la soppressione della risposta immunitaria provoca un'infezione cronica permanente invece della malattia letale (*si veda l'illustrazione in questa pagina*).

Sembra che una parte del processo che porta alla fatale malattia consista nell'attacco da parte del sistema immunitario dell'ospite contro il virus

Il virus CML, che provoca una coriomeningite linfocitaria, cioè un'infiammazione delle membrane del cervello, uccide il topo adulto quando viene iniettata direttamente in sito (*a*). Quando il virus CML viene invece iniettato in una vena, il topo adulto può morire rapidamente o manifestare una malattia meno grave seguita da immunità verso il virus (*b*). Nel corso della malattia letale le cellule infettate formano un gran numero di nuove particelle virali che infettano altre cellule. La risposta del sistema immunitario è un attacco sia contro le particelle virali, sia contro le cellule infettate. L'organismo muore quando un gran numero di cellule viene distrutto. Se l'infezione viene arrestata prima che riesca a interessare troppe cellule, l'animale sopravvive e mantiene l'immunità verso il virus. Quando la risposta immunitaria è debole, come nel topo neonato (*c*), l'iniezione del virus CML nel cervello provoca una infezione permanente. Il trattamento di topi adulti con raggi X o con droghe sopprime la risposta immunitaria; quando si iniettano in questi animali i virus CML, si manifesta una infezione permanente invece che la malattia letale (*d*). La risposta immunitaria dei topi trattati non è sufficiente a uccidere tutte le cellule infettate per cui esse restano per tutta la vita ammalate e continuando a formare virus antigeni virali. I complessi antigene e anticorpo che si formano sono trasportati dal sangue ai reni e provocano alterazioni gravi e progressive.

infettante e, probabilmente in grado anche maggiore, nell'attacco contro cellule infettate che portano proteine virali sulla loro superficie. Poiché il virus della CML per se stesso non è in grado di uccidere le cellule che infetta, non sorprende che la maggior parte delle cellule di un topo inoculato col virus della CML siano infettate ma continuino a funzionare finché vengono distrutte dalla risposta immunitaria dell'animale. Tutto ciò porta a una situazione paradossale in cui il sistema immunitario provoca una « malattia da virus » anche se l'infezione da virus è relativamente innocua.

Hotchin ha osservato che il topo permanentemente infettato non ha una vita di durata normale. I topi sono di solito in buona salute durante il primo anno dopo l'infezione, ma manifestano successivamente una malattia tra i cui sintomi vi è un danno renale simile a quello riscontrato in una malattia umana, il lupus eritematoso sistemico (LES). Michael B.A. Oldstone e Frank J. Dixon della Fondazione Scripps a La Jolla in California hanno dimostrato che i topi con una infezione permanente da virus CML erano affetti da una malattia da complesso immunitario. I topi producevano anticorpi che formavano complessi virus-anticorpi. Questi complessi venivano sequestrati nei capillari dei reni e davano luogo a una glomerulonefrite infiammatoria. Danni simili nei reni dei pazienti affetti da lupus eritematoso hanno indotto alla ricerca dei virus che possano essere responsabili della malattia. Oldstone e Dixon hanno anche trovato che il ceppo di topi Nuova Zelanda è geneticamente predisposto alla malattia da complesso immunitario quando viene infettato con certi virus.

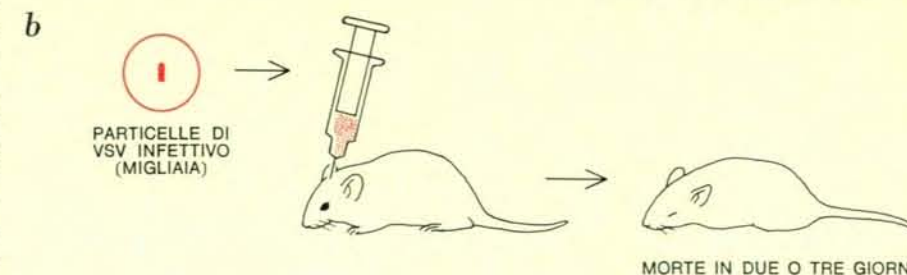
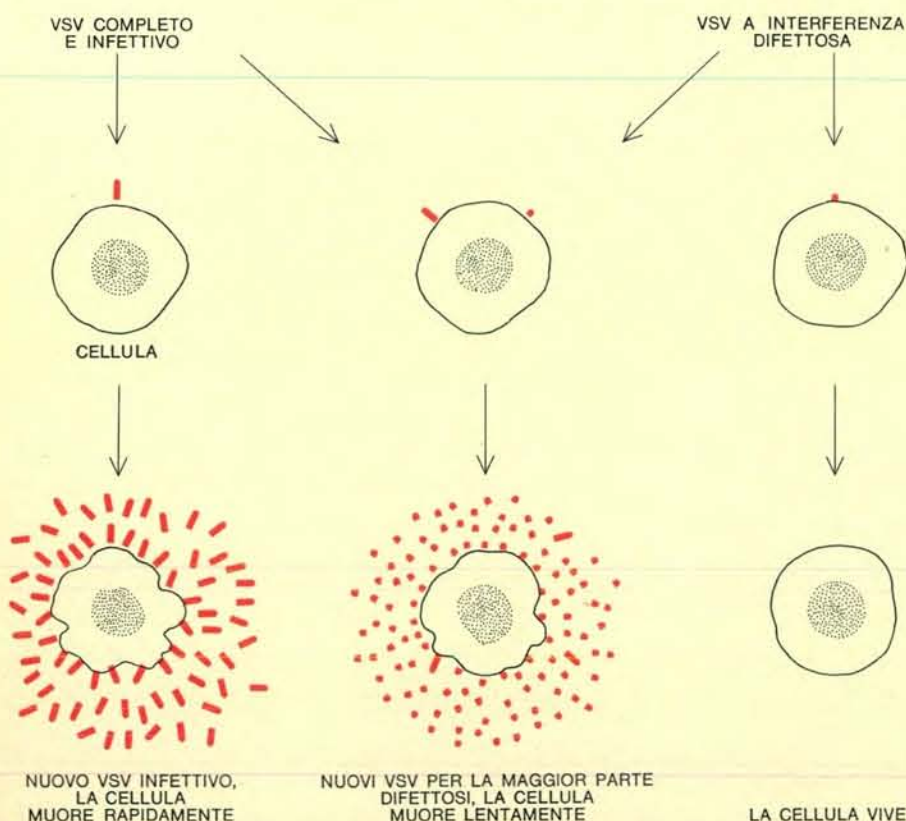
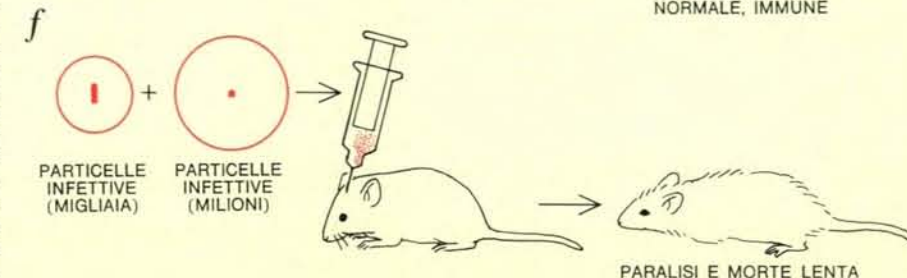
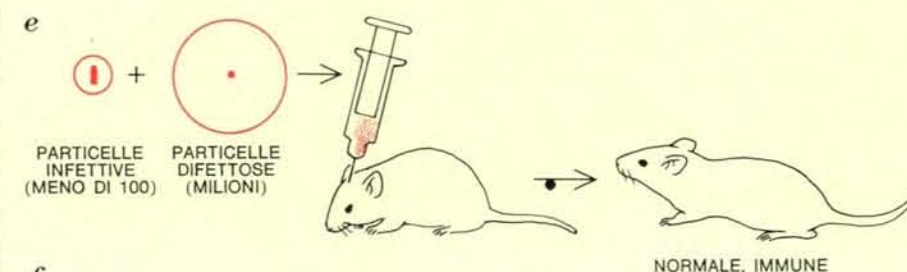
L'encefalopatia del visone, conosciuta anche col nome di malattia del visone aleutino, colpisce anch'essa animali geneticamente predisposti. Gli animali colpiti appartengono a un ceppo omozigote blu-grigio, cioè un ceppo in cui il gene per questo carattere viene ereditato da entrambi i genitori. La malattia provoca una eccessiva risposta immunitaria che fa aumentare i complessi circolanti di virus e anticorpi, i quali, a loro volta, provocano un massiccio danno tissutale.

In seguito alla scoperta che alcuni animali sono geneticamente predisposti a reagire troppo energicamente, troppo debolmente o in modo sbilanciato a infezioni virali permanenti, i ricercatori hanno cominciato a chiedersi se alcune malattie autoimmunitarie associate nell'uomo con l'invecchiamento, potessero anche implicare una infezione virale permanente. Alcuni ricercatori hanno riferito alcune prove preliminari che fanno pensare che i virus siano gli agenti causali dell'artrite reumatoide e di altre malattie degenerative; tuttavia mancano ancora prove definitive.

Uno degli aspetti piú interessanti dell'infezione virale è la capacità di alcuni virus di deprimere completamente o in parte il sistema immunitario in certe situazioni e di migliorare la risposta immunitaria in altre. Ciò può essere dovuto in parte al fatto che alcuni virus crescono in cellule, come per esempio i linfociti, che partecipano alla risposta immunitaria. Una grande varietà di virus (CML, tumori da RNA, morbillo, influenza e poliomielite) possono deprimere la produzione di anticorpi contro altri antigeni. Cosí nel corso di infezioni virali permanenti il virus stesso può iniziare e mantenere nel sistema immunitario aberrazioni permanenti, che a loro volta fanno sopravvivere il virus.

I virus del gruppo dell'*Herpes* si dimostrano particolarmente adatti a provocare infezioni permanenti o ricorrenti nell'uomo. L'*Herpes simplex* tipo 2 è la causa di ricorrenti lesioni genitali proprio come il tipo I è la causa ricorrente delle comuni vesciche delle labbra; si stanno accumulando varie prove che la ricorrente infezione genitale da *Herpes* possa essere la causa del carcinoma del collo dell'utero. Non sappiamo ancora come mai la maggior parte delle persone guarisca completamente dal virus che provoca la varicella (*Herpes zoster*), mentre poche persone ospitano il virus per tutta la vita e sono colpite periodicamente da dolorosi attacchi di fuochi di S. Antonio (lesioni varicellose localizzate lungo il decorso dei nervi sensitivi).

Un altro virus del gruppo dell'*Herpes*, il virus di Epstein-Barr, è stato fortemente sospettato di essere la causa del linfoma di Burkitt, un tumore del tessuto connettivo della mascella, e del carcinoma nasofaringeo (si veda l'articolo *Virus erpetici e cancro* di Keen A. Rafferty jr, in « Le Scienze » n. 65, gennaio 1974). Anche in questo caso i pazienti con mononucleosi infettiva formano anticorpi contro il virus di Epstein-Barr, ma non tutti coloro che sono infettati dal virus sono affetti da mononucleosi. Probabilmente il virus di Epstein-Barr può restare inattivo nell'organismo per anni. Esso di regola non provoca la morte delle cellule dove si riproduce, ma può trasformare le cellule in modo che in alcuni individui alla fine si sviluppa il cancro.

Le circostanze capaci di far variare un virus come quello del morbillo, che di solito uccide le cellule che infetta, in un virus permanente e non letale non sono conosciute. Alice S. Huang della Facoltà di medicina di Harvard e David Baltimore del Massachusetts Institute of Technology hanno formulato l'ipotesi che i virus a interferenza difettosa possano giocare un ruolo nell'instaurare una infezione virale permanente. Questi tipi di virus difettosi sono in realtà particelle virali imperfette che mancano di larghe porzioni di RNA o DNA nel loro genoma e sono per questo motivo incapaci di riprodursi; tuttavia essi possono riprodursi in cellule simultaneamente infettate con genomi virali intatti. Nel nostro laboratorio, Michael Doyle, Luis Villareal e io abbiamo dimostrato che

Colture di cellule infettate da particelle intatte di virus VSV formano migliaia di nuove particelle virali intatte attraverso un processo di gemmazione. Normalmente tutte le cellule infettate muoiono (*a sinistra*). L'aggiunta di VSV a interferenza incompleta non provoca la formazione di nuove particelle virali, perché il virus difettoso non possiede il codice genetico per la sintesi di tutte le necessarie proteine virali (*a destra*). Le cellule non muoiono se non quando milioni di particelle difettose vengono aggiunte per ciascuna cellula. Quando, sia il virus intatto, sia il virus difettoso infettano la stessa cellula, vengono formate molte nuove particelle sia difettose, sia infette e virulente (*al centro*). I virus difettosi utilizzano proteine sintetizzate dal virus intatto e in questo modo interagiscono con la velocità di produzione delle particelle virali intatte. Il grado di interferenza dipende dal numero delle particelle virali aggiunte e dal rapporto tra virus a interferenza difettosa e virus intatti. La doppia infezione è sempre letale.

I virus a interferenza difettosa non sono in grado da soli di provocare una malattia, ma possono modificare l'andamento di una malattia virale normalmente letale. Un topo, a cui vengano iniettate un piccolo numero (meno di 100) di particelle di virus completo e infettivo di VSV, muore in tre o quattro giorni (*a*). L'iniezione di un numero maggiore di particelle infettive accelera il processo fatale (*b*). Un topo a cui si iniettano poche particelle difettose VSV non mostra risposte apprezzabili (*c*). Quando invece vengono iniettate milioni di particelle virali difettose, non si ha infezione, ma si sviluppa uno stato immunitario verso il virus (*d*). L'abbondante dose di virus difettoso funziona come un vaccino: contiene una quantità di antigene sufficiente a stimolare la produzione di anticorpi. Quando una piccola, ma letale dose del virus infettivo viene iniettata assieme con una grande dose di virus difettoso, gli animali sono protetti, sopravvivono tutti all'infezione e diventano immuni (*e*). La riproduzione del virus infettivo è molto rallentata dal virus difettoso: ciò permette al sistema immunitario di arrestare l'infezione. Infine, il modello di una malattia virale lentamente progressiva si ritrova in topi a cui è stata iniettata un'abbondante dose di virus infettivo e un'abbondante dose di virus a interferenza difettosa (*f*). I topi sono permanentemente infettati e manifestano una paralisi progressiva degli arti posteriori. Essi peggiorano rapidamente e la maggior parte di essi muore entro due settimane. Ulteriori studi su come il virus difettoso interferisce con la riproduzione di virus normali potrebbero portare nuove informazioni sui meccanismi di altre infezioni virali lente.

virus a interferenza difettosa possono trasformare una infezione virale che normalmente porta in breve tempo a morte in una malattia piú lentamente progressiva. Quando il virus della stomatite vescicolare (VSV) viene iniettato nel cervello del topo, si sviluppa rapidamente una encefalite fatale.

Normalmente sono necessarie solo poche particelle di VSV per provocare la fatale malattia. Tuttavia, se si somministra ai topi una dose molto grande di VSV difettoso unita a una piccola dose di VSV intatto, gli animali diventano immuni e non vengono colpiti dalla malattia quando piú tardi a essi viene somministrata un'abbondante e letale dose di VSV intatto. Infine, quando una dose molto abbondante di VSV viene iniettata nel topo assieme col virus difettoso, gli animali non vengono colpiti da un'encefalite rapidamente fatale, ma da una malattia paralitica che logora l'organismo e porta alla morte dopo un periodo relativamente lungo (due settimane invece di due o tre giorni).

Abbiamo anche accertato che su colture di cellule il virus difettoso può far variare l'andamento dell'infezione da parte di virus VSV intatti. Normalmente tutte le cellule infettate dal virus VSV vengono distrutte, ma con l'aggiunta di virus difettoso si può ottenere una cultura cronicamente infettata portatrice del virus. Le cellule portatrici continuano a crescere producendo il virus VSV sia intatto, sia difettoso (*si veda la figura a pagina 32*). Durante la crescita nelle colture portatrici avvengono delle mutazioni nel virus e queste variazioni modificano la capacità del virus di provocare la malattia nel topo. Ulteriori studi sui virus a interferenza difettosa potrebbero chiarire considevolmente il meccanismo generale delle infezioni virali e permanenti.

Si ritiene che il virus a interferenza difettosa abbia azione competitiva nei confronti del virus intatto per ciò che riguarda il processo di riproduzione, esso infatti deve prendere a prestito alcune proteine formate dal virus intatto perché non contiene l'informazione genetica per la produzione di tutte le necessarie proteine virali. Questo meccanismo competitivo interferisce con la velocità di produzione dei virus intatti. Poiché i virus infettivi prodotti sono in minor numero, il processo infettivo è molto rallentato.

È noto che molti virus umani e animali producono particelle virali a interferenza difettosa. Questi virus difettosi giocano probabilmente un ruolo importante nell'evoluzione virale. Il virus a interferenza difettosa potrebbe impedire al virus intatto di manifestare la sua completa virulenza; in caso contrario il virus potrebbe decimare l'intera popolazione dell'organismo ospite, eliminando in questa maniera se stesso. L'occasionale e accidentale sviluppo di malattie da virus lenti in organismi ospiti con alcuni difetti del sistema immunitario potrebbe essere solamente un malaugurato effetto secondario dell'attività di un virus difettoso.

Virus a interferenza difettosa sono stati ritrovati anche in colture cellulari infettate con virus della coriomeningite. Ciò indica che i virus a interferenza difettosa protrebbero essere coinvolti in questa malattia del topo e in malattie autoimmunitarie e da complesso immunitario.

Ulteriori ricerche metteranno in luce probabilmente altri meccanismi molecolari per mezzo dei quali virus normalmente virulenti possono instaurare infezioni cellulari permanenti e non letali. È tuttavia già da ora chiaro che può essere ingannevole dare a un virus il nome della piú comune infezione acuta che esso provoca. Cosí facendo si tende a sviare l'attenzione dagli altrettanto importanti processi degenerativi che possono essere provocati dall'infezione virale cronica.

MALATTIE DI EZIOLOGIA VIRALE QUASI CERTA	SINTOMI TIPICI	PROVE
Creutzfeldt-Jacob	Funzioni mentali alterate, demenza presenile, esito normalmente letale	Agente trasmissibile dimostrato, ma non identificato
Kuru	Funzioni mentali alterate, demenza presenile, esito normalmente letale	Agente trasmissibile dimostrato, ma non identificato
MALATTIE PER LE QUALI L'EZIOLOGIA VIRALE È STATA IN PARTE DIMOSTRATA	SINTOMI TIPICI	PROVE
Linfoma di Burkitt Carcinoma nasofaringeo	Tumori maligni	Coinvolto il virus di Epstein-Barr
Carcinoma del collo dell'utero	Tumori maligni	Coinvolto il virus dell'Herpes simplex
Leucemie	Cellule del sangue anomale	Trovati negli animali virus RNA che causano tumori; prove preliminari nell'uomo
Lupus eritematoso	Infiammazione cutanea	Coinvolti paramixovirus
Sclerosi multipla	Perdita della coordinazione muscolare	Coinvolti il virus del morbillo e della parainfluenza
Leucoencefalopatia progressiva multifocolare	Sonnolenza, andatura incerta, visione offuscata, porta di norma a morte	Isolato il virus del tipo SV 40
Panencefalite sclerosante sub-acuta	Funzioni mentali e motorie danneggiate, porta di norma a morte	Coinvolti il virus del morbillo e in qualche caso i virus del tipo SV 40
ALTRE MALATTIE DI POSSIBILE EZIOLOGIA VIRALE	SINTOMI TIPICI	PROVE
Sclerosi laterale amiotrofica	Atrofia muscolare, porta di norma a morte	Informazioni preliminari sparse; nessuna prova concreta finora
Diabete	Perdita di cellule insulari pancreatiche, mancanza di insulina, iperglicemia	
Glomerulonefrite	Infiammazione dei reni	
Anemie emolitiche	Distruzione degli eritrociti	
Distrofia muscolare	Atrofia muscolare progressiva	
Neurite del nervo ottico	Danni del nervo ottico fino alla cecità	
Morbo di Parkinson	Tremori	
Psoriasi	Desquamazione	
Artrite reumatoide	Infiammazione delle articolazioni	
Colite ulcero-emorragica	Ulcerazione cronica e ricorrente del colon	

Le malattie degenerative sopraelencate sono alcune tra le malattie umane che meritano uno studio completo prima che si possa escludere la loro origine virale. Tuttavia, in queste malattie, l'isolamento di un virus non prova necessariamente che la malattia sia provocata da questo virus che potrebbe essere presente per caso o per un indebolimento del sistema immunitario.

La liofilizzazione

È un processo fisico sempre piú frequentemente utilizzato dalla tecnologia moderna per la perfetta conservazione a lungo termine di varie sostanze e, in particolare, di farmaci e alimenti

di Sonia Martinis Marchi Jellicich

La liofilizzazione è un metodo di essiccamento di prodotti o sostanze preventivamente congelate in cui la componente liquida, solidificata, viene eliminata sotto vuoto, pressoché totalmente, sotto forma di vapore, mentre il prodotto cosí essiccato risulta estremamente reidratabile. Gli anglosassoni usano anche la terminologia «*Freezing and Drying by Sublimation in Vacuum*» o semplicemente «*Freeze Drying*». È chiaro che il termine italiano si riferisce alla caratteristica finale delle sostanze cosí trattate cioè alla loro affinità per il solvente, mentre i termini inglesi al procedimento stesso. Risalire all'origine di questa tecnica è ben difficile anche se, in genere, si accettano come originali gli studi di Bordas e d'Arsonval del 1906 e quelli del 1909 di Shackell, ma, già agli inizi del secolo XIX, alcuni ricercatori avevano compiuto ricerche in questo campo. Le ricerche ripresero con nuovo vigore tra il 1935 e il 1940 in particolare nel campo della conservazione dei microrganismi. Approfonditi e sperimentati i vantaggi che poteva offrire la liofilizzazione, la si applicò, nel corso dell'ultimo conflitto mondiale per la conservazione del plasma umano. Il plasma può essere agevolmente conservato sia allo stato liquido sia congelato, ma richiede per l'immagazzinamento temperature ben definite e rigorose e, nel caso del prodotto congelato, una tecnica ben precisa per lo scongelamento. Parametri, questi, non facilmente rispettabili in periodo bellico e del tutto superflui per il plasma liofilo.

Il plasma liofilizzato richiede per una buona conservazione che assicuri una validità di 5 anni, un tasso d'umidità residua inferiore allo 0,5 per cento, la chiusura sotto vuoto o sotto azoto (parametri ottenibili nel corso e alla fine del processo) e una temperatura esterna di +20°C. La tecnica è talmente corrispondente che si è estesa anche ai derivati plasmatici (crioprecipitati, antiemofilici, albumina umana, fibrinogeno ed altri).

Caratteristica specifica delle sostanze liofilizzate, oltre a quella di conservarsi inalterate nel tempo, è quella della igroscopicità per cui esse possono essere ricostituite prontamente con l'aggiunta di un adeguato liquido. Lo spiccato potere igroscopico è conseguente alla sublimazione dei cristalli di ghiaccio che, fuoriuscendo, lasciano nella massa liofilizzata una struttura porosissima. Il processo in effetti si identifica con quello fisico della «sublimazione»: passaggio diretto dallo stato solido a quello di vapore; ma, in questo caso, con somministrazione di calore (calore latente di sublimazione) e sotto vuoto.

Nei prodotti liofilizzati si riscontra la stasi dei processi chimici e biologici, l'igroscopicità, la conservazione delle caratteristiche organolettiche (nel caso degli alimentari), del volume iniziale e una sensibile decolorazione che perdura sino all'atto della ricostituzione del prodotto con acqua. Il fenomeno è generalizzato per tutti i liofilizzati, anche per i farmaci, come si può rilevare dalle figure nella pagina a fronte. La decolorazione si determina nella fase del congelamento in quanto essa è collegata alla forma cristallina della massa: un ulteriore schiarimento si ha all'atto della sublimazione dei cristalli.

La variazione del colore delle specialità farmaceutiche liofilizzate all'atto della solubilizzazione, con solvente incolore, non comporta prevenzioni psicologiche da parte del consumatore. Il farmaco, nel concetto corrente, è tanto piú efficace quanto esso è misterioso mentre, per esempio, per gli alimentari si pretende la massima genuinità e la perfetta corrispondenza al prodotto originale, anche quindi nel colore e questo, non solo all'atto della ricostituzione, ma nel prodotto essiccato. Forse questo è uno dei fattori che frenano l'affermazione del liofilizzato sul mercato alimentare in cui è determinante anche l'aspetto esteriore del prodotto, oltre alle caratteristiche organolettiche e nutrizionali. D'altra parte è comprensibile, in un'epoca di sofisticazione dilagante, l'incertezza del consumatore; ma è certo che proprio il prodotto liofilizzato è quello che meno si presta a essere sofisticato specie se viene liofilizzato in pezzi come si può fare con la carne, i latticini, il pesce, ecc. Il liofilizzato è un prodotto «statico» in cui temporaneamente è stata bloccata ogni forma di attività, ma questa è latente e può essere ripristinata con opportuno solvente. Per il tecnico il ripristino pressoché totale del colore originale, dopo la ricostituzione del prodotto liofilizzato, è una delle visibili garanzie della rispondenza qualitativa del prodotto stesso.

L'applicazione industriale della liofilizzazione, in campo farmaceutico, iniziò con l'avvento degli antibiotici per estendersi poi ad altri medicamenti. Nella ricerca questa tecnica va sempre piú affermandosi e in campi talora impensabili quali quello nucleare per la concentrazione di sali radioattivi, nelle emulsioni per uso fotografico, nella medicina dei trapianti, nei reperti medico-legali. Le attuali e le possibili applicazioni future sono talmente numerose che praticamente è impossibile elencarle.

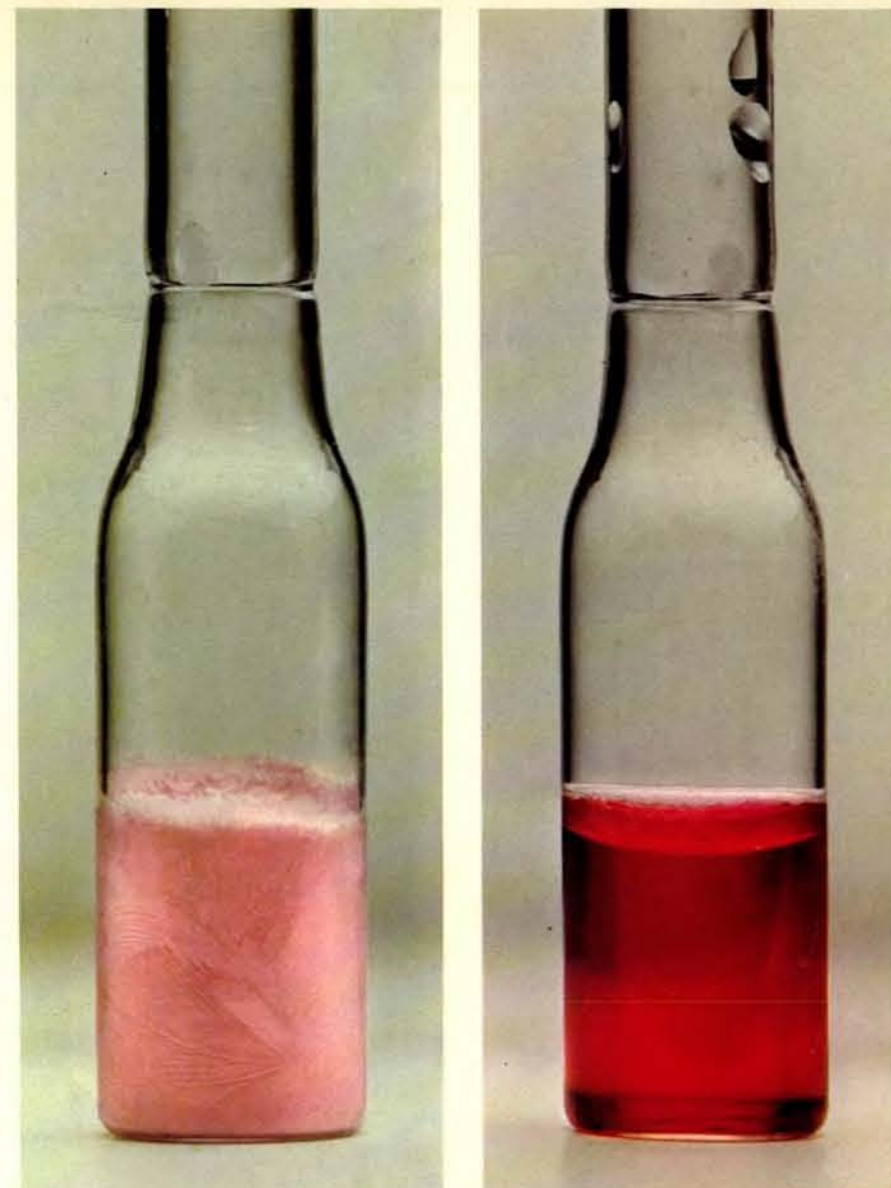
Il trionfo divulgativo, se cosí si può dire, della liofilizzazione nella società dei consumi fu decretato dai voli spaziali quando, seguendo ogni minimo particolare della vita degli astronauti, si venne a sapere che i pasti erano essenzialmente costituiti da cibi liofilizzati.

Ma quali sono i motivi validi, indipendentemente dall'astronautica, che hanno determinato e che determinano tuttora il crescente sviluppo della tecnica della liofilizzazione in rapporto agli altri sistemi di essiccamento e, quindi, di conservazione?

Rispetto agli altri metodi, che possono anche avvenire sotto vuoto, essa presenta il vantaggio enorme che l'essiccamento si effettua sí per somministrazione di calore, ma mantenendo la temperatura delle sostanze sempre a valore negativo e inferiore a quello del punto finale di congelamento, detto eutettico o zona d'eutessia; cosí operando non si interferisce sulle caratteristiche fisiche e chimiche dei prodotti che rimangono pressoché stabili.

Successivamente la temperatura dei prodotti, eliminata la componente liquida solidificata, è portata a valori positivi senza interferire sulla qualità finale e serve solo per ridurre l'umidità residua ai valori prestabiliti, specifici per ogni tipo di prodotto, per una buona conservazione nel tempo. Il processo di liofilizzazione può quindi in pratica essere definito come un processo di essiccamento a freddo. Oltre alle affascinanti e concrete possibilità offerte ai ricercatori il processo di liofilizzazione ha una vasta applicazione industriale nel campo farmaceutico e un graduale sviluppo in quello para-farmaceutico (prodotti dietetici) e alimentare.

La causa di questa differenziazione è da ricercarsi in un motivo molto semplice: quello economico. Il processo è costoso non solo per l'investimento delle macchine, ma principalmente per i consumi d'esercizio.

È intuitivo che nel trattamento dei farmaci l'incidenza del costo può essere ampiamente giustificata e assorbita se con questa tecnologia si possono risolvere problemi altrimenti irrisolvibili.

Specialità farmaceutica contenente vitamine del complesso B. Nella fiala liofilizzata di sinistra il contenuto ha una colorazione debolmente rosata, mentre nella fiala di destra, dopo l'aggiunta di un solvente incolore (acqua distillata), il prodotto assume un colore rosso rubino che è uguale a quello che aveva prima del trattamento e caratteristico delle soluzioni contenenti vitamina B_{12} alla concentrazione dell'1 per mille.

Anche gli alimenti vengono ormai sempre piú di frequente liofilizzati con successo; il procedimento è conveniente soprattutto nel caso di prodotti pregiati come il fungo porcino qui fotografato. A sinistra si può vedere un pezzo di prodotto liofilizzato dal colore sbiadito, mentre a destra, dopo reidratazione, il colore è tornato identico a quello dei funghi freschi.

La medicina, e di pari passo i medicamenti, sono andati sempre piú evolvendosi nei secoli e in particolare negli ultimi 50 anni. Il famoso, frutto della scienza galenica, approntato nel retro della farmacia per il singolo ammalato, con esigenze limitate di conservazione e di semplice somministrazione, si è trasformato in un prodotto industriale complesso derivante a sua volta da una maggiore disponibilità di principi attivi ottenuti con le piú svariate e piú progredite tecnologie di sintesi, di fermentazione, di estrazione ecc.

Questa evoluzione ha comportato nuove possibilità, nuovi concetti e nuove esigenze nella produzione del farmaco tra cui, fondamentale, la stabilità nel tempo delle sostanze medicamentose. La tecnologia della liofilizzazione è particolarmente adatta per risolvere i problemi inerenti la stabilità, l'associazione e la conseguente sicura conservazione di molti principi attivi instabili in soluzione o anche in sospensione. Essa esplica una duplice funzione: quella scientifico-terapeutica e quella economica. Scientifico-terapeutica in quanto rende utilizzabile una maggiore gamma di principi attivi; economica perché ne prolunga la validità, quindi la conservazione nel tempo.

Nei calcoli dei costi industriali una voce onerosa è quella rappresentata dai rientri dei farmaci di breve validità che, comunque, limitano l'entità della produzione in serie, incidendo direttamente sulla valutazione del costo globale del prodotto. Se di un farmaco si riesce a protrarre la scadenza, per esempio da sei mesi a due anni, è intuitivo che il vantaggio economico sarà tale da coprire del tutto quello derivante dal maggior costo del trattamento.

In genere l'arco produttivo di un farmaco sterile completato dai controlli chimico-biologici, dal confezionamento e dall'immissione alla catena di distribuzione può essere valutato, in media, in 30 giorni a cui, comunque, devono aggiungersi quelli di immagazzinamento prima della vendita finale. Quanto piú è breve la scadenza tanto piú è alta la percentuale di ritorno del farmaco al produttore che è impegnato a sostituirlo, gratuitamente, con altro di recente produzione. Non si sottovaluti, inoltre, la forzata rinuncia alle vendite di questi prodotti sui mercati esteri.

È chiaro quindi che la liofilizzazione in questi casi pur essendo costosa può essere piú che conveniente, tanto piú se non è sostituibile con altri sistemi. Inoltre, tra il consumo di specialità farmaceutiche e dei prodotti alimentari esiste una sostanziale differenza: le prime sono scelte a discrezione del medico curante che le impone, sotto la propria responsabilità, all'ammalato; i secondi, sono a totale e libero arbitrio del consumatore, che deve quindi essere in grado di giudicare e apprezzare le qualità dei prodotti e, in base a questi, accettare un prezzo maggiore rispetto agli stessi prodotti conservati con altri sistemi.

Si può realisticamente affermare che se la liofilizzazione ha avuto un lento avvio nell'industria conserviera, ciò è dovuto principalmente a errori di valutazione sia tecnica che economica. Con il passare degli anni questi errori si sono evidenziati, per cui oggi assistiamo a una razionale e progressiva utilizzazione di questa tecnologia con criteri non piú artigianali e avventuristici, ma industriali a cui dovrà fare riscontro una vasta campagna promozionale per far conoscere quali sono effettivamente i pregi e le garanzie dei prodotti liofilizzati.

Prima di illustrare la realizzazione di un apparecchio liofilizzatore è bene soffermarsi sulle specifiche funzioni richieste alle varie stazioni di lavoro di cui esso si compone. Chiarito per sommi capi il concetto fisico del metodo, facilmente si deduce che tre sono i fondamentali parametri che ne consentono l'effettuazione: il freddo, il calore e il vuoto (o bassa pressione), parametri che comportano un notevole e ben determinabile consumo d'energia.

Il freddo è coinvolto in tutte le tre fasi, in cui normalmente viene suddiviso il processo: congelamento, sublimazione, disidratazione finale.

Nella prima fase ha la funzione di sottrarre il calore di solidificazione dei prodotti e portare il livello termico a un valore inferiore a quello corrispondente al totale congelamento della massa. Nella seconda e terza fase, serve a trattenere su apposite superfici refrigerate i vapori che, a contatto con queste, si solidificano.

Nella figura sono schematizzati due diversi metodi per liofilizzare le sostanze. Nel primo caso la fiala viene posta direttamente sulla piastra termica portaprodotto, la massa congelata che essa contiene inizia a sublimare (cioè a passare direttamente allo stato di vapore) dalla superficie esterna libera dopo di che il fronte della sublimazione va man mano avanzando verso l'interno. Nel caso di una massa sospesa tra due piastre termiche, i fronti in cui inizia la sublimazione sono invece due. Il prodotto liofilizzato è chiaramente riconoscibile per il diverso colore che assume.

Il complesso destinato all'intrappolamento, in fase solida, dei vapori è genericamente detto « condensatore » ma ha anche la funzione di « crio-pompa ».

Il calore è coinvolto nella seconda e terza fase e rappresenta l'energia indispensabile per il passaggio di stato, da solido a vapore, e per la disidratazione finale della massa.

Il calore che viene fornito nella seconda fase non è altro che il « calore latente di sublimazione » ovvero la quantità di calore necessaria per sublimare l'unità di massa a temperatura costante e a pressione stabilita.

Le soluzioni tecniche adottate per il riscaldamento possono essere molteplici; la trasmissione del calore al prodotto può realizzarsi mediante tre tradizionali sistemi: conduzione, irraggiamento, convezione, spesso associati tra loro. Qualsiasi sia però il metodo previsto, che a sua volta condiziona il sistema di carico dei materiali, dovrà essere tale da garantire al massimo l'omogeneità e l'intensità del riscaldamento a tutto il carico con sistemi esatti di controllo e di emergenza.

Bisogna tenere presente infatti che nella realizzazione impiantistica devono ridursi al minimo le dispersioni termiche per limitare i consumi alle reali necessità del ciclo.

Il vuoto, o piú precisamente la bassa pressione, è anch'esso coinvolto nella seconda e terza fase: l'azione dei gruppi di pompaggio, che diminuisce la concentrazione gassosa in un determinato ambiente, è indispensabile per mantenere la pressione parziale del vapore a un valore inferiore alla tensione del vapore saturo del solvente del prodotto. Il gruppo di pompaggio con la sua azione aspirante è in grado di eliminare inizialmente i gas permanenti e, successivamente, gli incondensabili (inglobati) e i condensabili che emanano dal prodotto direttamente verso l'esterno del sistema o piú comunemente verso il condensatore in grado di intrappolarli.

La totale e diretta eliminazione dei vapori condensabili tramite il gruppo di pompaggio può essere realizzata negli impianti da laboratorio con l'eventuale interposizione di una trappola chimica mentre, su scala industriale, è piú conveniente utilizzare il « condensatore » dati gli enormi volumi dei vapori condensabili conseguenti alla pressione di regime e questo sia nel caso della liofilizzazione farmaceutica che in quella alimentare.

È necessario tenere presente infatti che il volume specifico dei vapori aumenta notevolmente con il diminuire della pressione: 1 chilogrammo di acqua allo stato di vapore, alla pressione di 1 torr ha un volume specifico di 1000 m³ e a 0,1 torr di 10 000 m³. Il processo di liofilizzazione, che viene effettuato a medio e ad alto vuoto, cioè a pressione di regime variabile rispettivamente tra quella atmosferica e 1 torr e tra 1 e 10^{-3}torr, dà luogo quindi alla produzione di grandi volumi di vapori, soprattutto se si ha a che fare con forti quantità di sostanze da trattare.

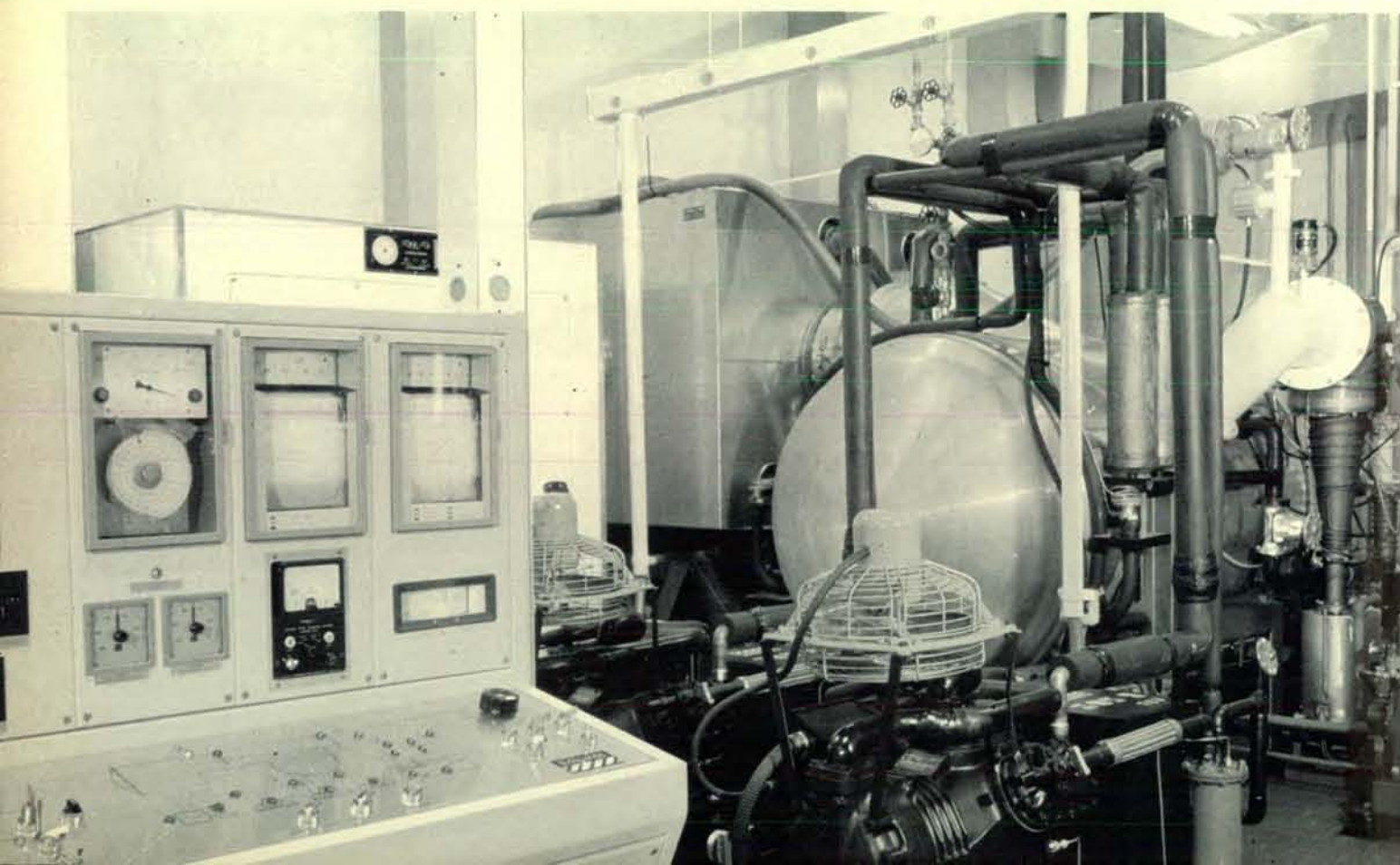
La liofilizzabilità e la durata dei cicli produttivi di un prodotto possono essere condizionate dalle sue caratteristiche chimico-fisiche, ma non vincolate dalle prestazioni degli impianti. Nella progettazione e realizzazione si devono quindi creare i presupposti meccanici, geometrici e funzionali che consentano l'ottimale rendimento di tutti i componenti per cui sia garantito lo scambio energetico con i prodotti in lavorazione. Per la realizzazione di questi presupposti deve esistere una profonda conoscenza impiantistica che coinvolga problemi di termodinamica, di vuoto, di meccanica. Concettualmente i liofilizzatori industriali, destinati a qualsiasi tipo di produzione, sono molto simili, variano solo alcuni componenti e proprio in funzione delle caratteristiche dei prodotti in lavorazione.

Principalmente differiscono nelle temperature e nelle pressioni di regime, nel congelamento dei prodotti – che può essere esterno o interno al liofilizzatore – e nei sistemi di carico all'interno di questo e, quindi, nel metodo di riscaldamento. Nella panoramica che seguirà, relativa a un liofilizzatore tipo, verranno focalizzate di volta in volta le differenze operative.

L'autoclave, detta anche « camera di essiccazione », è di norma metallica, di forma parallelepipeda o cilindrica: la scelta della geometria viene valutata in funzione dei fattori sia economici che pratici che coinvolgono il massimo sfruttamento del volume interno, l'uniformità di carico e la manovrabilità interna. Il materiale utilizzato per la costruzione non deve essere ossidabile per cui si ricorre all'acciaio inox o anche a materiale meno pregiato, ma adeguatamente trattato per resistere alle corrosioni. L'autoclave e, quindi, i relativi passanti d'adduzione dei fluidi refrigeranti e riscaldanti devono essere a perfetta tenuta di vuoto, in grado di sopportare la differenza di pressione esistente tra l'interno e l'esterno di essi e che è di una atmosfera.

Se oltre alla tenuta al vuoto si rende necessaria quella a pressione, per permettere la sterilizzazione interna con vapore d'acqua saturo e sotto pressione, valgono i parametri di resistenza e sicurezza delle comuni autoclavi di sterilizzazione a vapore. L'autoclave può essere a una o due porte, passante tra la sala macchine e la sala di carico e scarico dei materiali (*si vedano le foto a pagina 40 e lo schema a pagina 42*).

Il disegno rappresenta lo schema tecnologico di un liofilizzatore industriale in cui possiamo riconoscere vari elementi. A sinistra in alto si vede l'autoclave in cui viene fatto il vuoto per facilitare la sublimazione dei prodotti preventivamente congelati mediante la refrigerazione delle piastre portaprodotto (*la linea refrigerante è rappresentata in colore scuro*). Mettendo in funzione il sistema di riscaldamento delle piastre (*colore chiaro*), si ha il rapido passaggio della parte acquosa solidificata delle sostanze allo stato di vapore, che viene immediatamente incanalato e trattenuto entro l'apposito condensatore e successivamente eliminato mediante il sistema di pompaggio.

Veduta parziale di una sala sterile per la liofilizzazione di prodotti farmaceutici. A destra è chiaramente visibile la porta dell'autoclave del liofilizzatore e sulla parete di fondo l'armadio per la conservazione sotto vuoto. Le zone in cui avviene il carico e principalmente lo scarico del materiale trattato devono essere rigorosamente sterili e avere un basso tasso di umidità relativa in modo da evitare la reidratazione dei prodotti. Nella fotografia in basso si può vedere la sala macchine dello stesso reparto di liofilizzazione. In primo piano il tavolo di comando del liofilizzatore in cui è incorporata la strumentazione di controllo.

Nella produzione farmaceutica la zona di carico e scarico del prodotto è unica e normalmente fa parte della zona sterile in quanto i prodotti devono essere trattati asetticamente; in questo caso, l'autoclave è a porta unica e passante tra la sala macchine (non sterile) e il locale adiacente sterile.

Nella produzione alimentare la zona di carico del prodotto entrante e di scarico del prodotto finito possono essere distinti, per cui l'autoclave è a doppia e contrapposta apertura. Il locale destinato all'estrazione del materiale liofilizzato deve essere mantenuto a un tasso, prestabilito e costante, d'umidità relativa per evitare la reidratazione del prodotto.

Nell'autoclave è sistemato il complesso delle piastre termiche che, oltre a esplicare la specifica funzione di scambio energetico, possono essere base di appoggio diretto per il materiale, in questo caso, vengono anche definite piastre-portaprodotto.

Nei liofilizzatori destinati alla produzione farmaceutica le piastre hanno la duplice funzione, termica e di sostegno, oltre a essere riscaldabili sono refrigerate per cui è possibile effettuare con continuità, all'interno di esse, tutte e tre le fasi del processo: congelamento - sublimazione - essiccamento.

Le piastre sono di norma fisse ma possono essere rese mobili, in senso verticale, mediante opportuni sistemi idraulici. Questa possibilità di movimento è richiesta per consentire, prima dell'estrazione finale, la chiusura sotto vuoto o gas inerte di flaconcini contenenti prodotti altamente ossidabili e igroscopici.

Per effettuare la chiusura dei flaconcini è anche possibile ricorrere a sistemi piú economici di quanto non sia quello idraulico prevedendo sopra ogni piastra portaprodotto, fissa, una contropiastra sospesa ma resa mobile con un semplice sistema pneumatico.

In precedenza si è affermato che i tre fattori determinanti del processo di liofilizzazione sono: il freddo, il caldo, il vuoto; vale a dire che il processo è condizionato dallo scambio termico tra i prodotti e gli elementi « energetici » (in questo caso le piastre) e dalla conseguente possibilità di evaporazione della fase solida dall'interno di essi.

Per le piastre si richiedono gli accorgimenti costruttivi e termodinamici in grado di garantire al massimo un uniforme rendimento termico sia nella fase congelante che in quella riscaldante.

Ciò allo scopo di ottenere una omogeneità, ma anche di contenere al massimo la durata dei cicli.

Le piastre sostanzialmente sono refrigerabili con due sistemi; il primo può essere definito « diretto » in quanto il fluido frigorigeno viene fatto espandere direttamente nelle piastre in adatte canalizzazioni. Il secondo, « indiretto », viene attuato con la circolazione di un fluido diatermico che viene raffreddato in uno scambiatore di calore separato in cui si fa espandere il fluido frigorigeno.

La scelta tra questi due sistemi è determinata da particolari esigenze dei prodotti o anche dalla eventuale convenienza di realizzare un'unica centrale del freddo da sfruttarsi in tempi successivi su piú di un liofilizzatore.

In linea di massima, non considerando la « centralizzazione », il sistema refrigerante « indiretto » è piú oneroso in quanto comporta un maggiore costo dell'impianto e maggiori consumi di esercizio.

I tempi di congelamento e le temperature finali dei prodotti condizionano la scelta dei compressori meccanici e del fluido refrigerante.

Il riscaldamento delle piastre termiche è normalmente realizzato per circolazione di un fluido diatermico riscaldato elettricamente o mediante vapore in uno scambiatore di calore separato; è richiesta una perfetta regolazione non solo della circolazione all'interno delle piastre, (si ricorre a circuiti interni a labirinto o a serpentine in controcorrente), ma anche della temperatura che deve mantenersi a livelli ben prestabiliti.

Nei liofilizzatori alimentari le piastre hanno principalmente una funzione termica positiva. Sono sistemate a « pettine » per consentire, tramite sistemi carrellati centrali, il posizionamento dei vassoi contenenti il materiale (congelato all'esterno in una stazione refrigerante) in posizione fissa ed equidistante tra di esse. Il complesso delle piastre riscaldanti può essere reso mobile per consentire, a seconda delle necessità, l'avvicinamento o meno al materiale stesso. Il congelamento viene effettuato, come detto, all'esterno del liofilizzatore dati gli enormi quantitativi dei prodotti da trattare e i problemi dell'approvvigionamento stagionale e relativa conservazione. Senza contare che, specie per i prodotti liquidi o semi-liquidi, vi sono delle necessità di pretrattamento per ridurre la percentuale acquosa da liofilizzare. Queste operazioni consistono nella preconcentrazione; si è dimostrata particolarmente utile, allo scopo di non deteriorare le caratteristiche organolettiche, la crioconcentrazione in cui il prodotto congelato viene ridotto in granuli del diametro di 2-4 mm per aumentare la superficie evaporante e ridurre lo spessore del prodotto. Questi fattori possono influire direttamente sulla velocità della sublimazione e quindi sulla durata dei cicli di liofilizzazione.

Uno dei limiti alla velocità di essiccamento intrinsechi al prodotto, oltre a quello delle caratteristiche chimico-fisiche, è dato dalla resistenza opposta dagli strati essiccati al passaggio dei vapori che defluiscono dall'interno della massa. È chiaro che tanto piú è ridotto lo spessore di questo strato tanto inferiori sono le difficoltà che incontrerà il vapore a emergere: per cui si tende a limitare lo spessore dei materiali da liofilizzare.

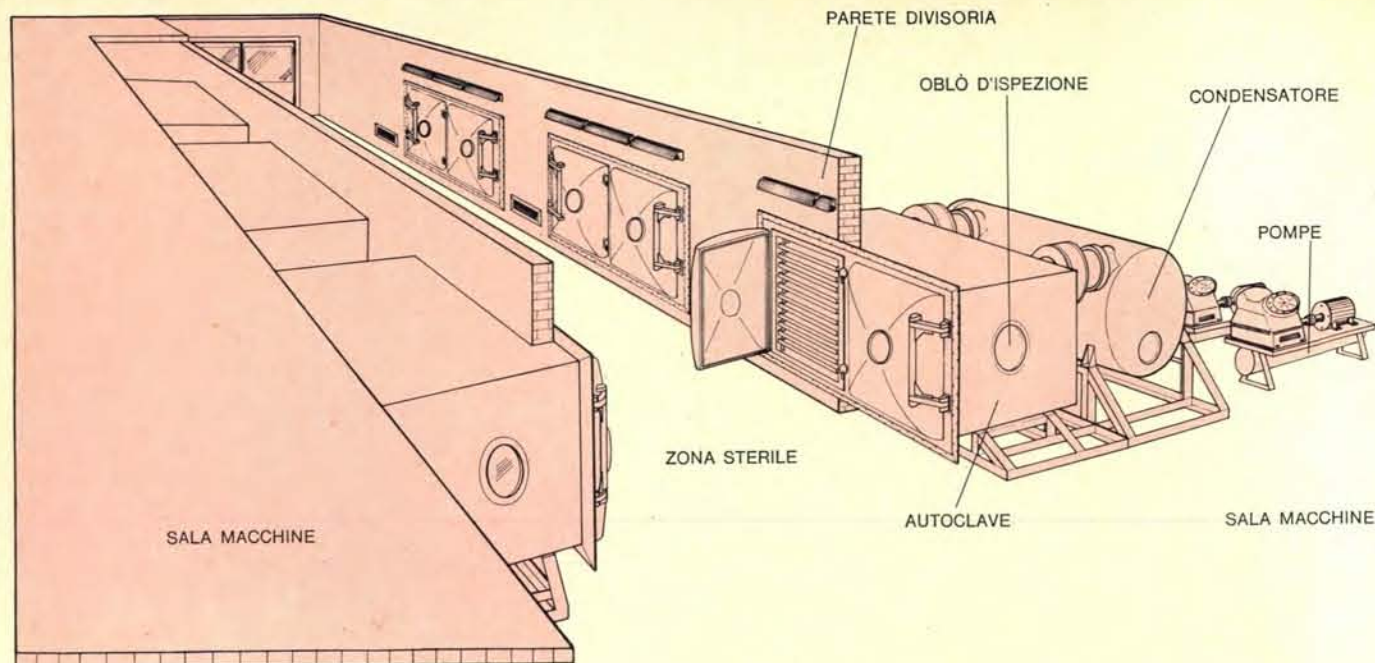
Il riscaldamento delle piastre, normalmente annerite per incrementarne l'irraggiamento termico, può avvenire anche tramite vapore. I criteri costruttivi di queste piastre sono meno impegnativi di quelle destinate a essere anche « portaprodotto ».

Il condensatore o meglio l'involucro contenente le superfici refrigerate e quindi condensanti, a perfetta tenuta di vuoto, è di norma esterno all'autoclave e a essa collegato tramite un condotto a sezione calcolata per il regolare deflusso dei vapori; una valvola di intercettazione ne consente l'isolamento e anche, se richiesto da particolari accorgimenti operativi, la riduzione del flusso dei vapori. L'indipendenza delle due zone è richiesta per consentire l'autonomia di funzionamento delle singole sezioni nelle interfasi preparatorie e anche nella prima fase del processo.

Le « interfasi » tra un ciclo e l'altro sono rappresentate dalle operazioni di carico e scarico e, se previsto, di congelamento dei prodotti nell'autoclave, dallo scongelamento del ghiaccio depositatosi sulle superfici refrigerate del condensatore e dal ripristino delle temperature di regime delle stesse; tutte condizioni indispensabili per dare corso alla successiva lavorazione.

Questa autonomia comporta una razionale successione operativa abolendo completamente i tempi morti.

Le superfici refrigerate, come l'involucro, sono costruite in metallo resistente all'attacco chimico dei vapori. Vengono normalmente utilizzate delle serpentine anche alettate per aumentare la superficie d'impatto, o delle piastre e serpentine interne. La refrigerazione delle superfici viene ottenuta mediante fluido refrigerante in espansione, in allagamento e per circolazione forzata.

Il condensatore, sempre interposto tra l'autoclave (o ambiente sublimante) e la stazione pompante, esplica una funzione di « barriera » in quanto trattiene, condensandoli, i vapori riservando

Schema di una serie di impianti multipli di liofilizzazione. Al centro la zona sterile per il carico e scarico del materiale su cui si aprono i portelli delle autoclavi. Ai lati, le sale macchine, isolate mediante una parete dalla zona centrale.

al sistema di pompaggio la sola eliminazione degli incondensabili.

L'efficienza del sistema condensante è determinata da vari fattori concomitanti: dalla disposizione geometrica delle superfici refrigerate che deve favorire l'impatto diretto del vapore, dalla potenzialità frigorifera e dalla realizzazione del circuito relativo.

La potenzialità frigorifera viene calcolata per condensare la massa di vapori che si svilupperanno, mediamente, nel corso di un prefissato ciclo di sublimazione. In base a questa specifica richiesta, considerando anche la temperatura cui dovranno essere portate le superfici refrigerate per condensare i vapori, il livello di temperatura è imposto dalle caratteristiche fisiche dei prodotti. Per esempio per condensare 1 chilogrammo di acqua sono necessarie 640 chilocalorie per ora (valore teorico); se si prevede di evaporare 10 chilogrammi di acqua e di mantenere il livello di temperatura di condensazione a –50 °C deve essere previsto un gruppo frigorifero in grado di erogare una potenza di 6400 Kcal/h a –55 °C. Queste rese devono essere valutate alla effettiva temperatura di espansione del compressore, in quanto nella realizzazione di un circuito frigorifero immancabilmente si hanno delle perdite di carico (contenute, normalmente, nell'ordine del 10 per cento). Non si dimentichi che la differenza di temperatura tra il prodotto e la superficie condensante e la conseguente differenza di pressione hanno anche una diretta influenza sull'evaporazione, tanto piú sono alte queste differenze quanto piú si favorisce l'azione « aspirante » della superficie refrigerata; non per niente al sistema condensante è riconosciuta la funzione di « criopompa ».

Le temperature delle superfici del condensatore oscillano, di media, tra i –30 °C e –50 °C per i prodotti alimentari, tra i –40 °C e –80 °C per quelli chimico-farmaceutici; temperature inferiori a questi valori, almeno per ora, non sono di interesse industriale perché eccessivamente costose. Studi molto interessanti sono in corso per l'uso di gas liquefatti quali l'azoto: è auspicabile che siano coronati da successo in quanto si eliminerebbe la complessità delle attuali linee frigorifere anche se in effetti la tecnica, in questo campo, è molto progredita e offre varie possibilità di utilizzo anche alle bassissime temperature.

Il gruppo di pompaggio è direttamente collegato, ma isolabile, al condensatore e, tramite esso, all'autoclave; viene chiamato in causa quando si dà inizio alla seconda fase del processo: quella sublimante.

La scelta del gruppo viene fatta in base ai calcoli volumetrici del complesso da evacuare, al tempo necessario per portarlo al grado di pressione necessaria per l'inizio della sublimazione. Il gruppo deve essere in grado di eliminare i gas permanenti, iniziali e quelli inglobati nella massa del prodotto e compensare nel contempo le inevitabili perdite del sistema. L'azione pompante diminuendo la concentrazione gassosa che circonda la massa, favorisce il deflusso dei vapori dall'interno di essa, valido ausilio dell'energia fornita al prodotto dal sistema riscaldante. Le pompe sono di tipo rotativo a bagno d'olio, a singolo o doppio stadio, spesso combinate con pompe booster meccaniche o a vapore.

Per la lettura delle pressioni sono preferiti i vacuometri elettrici tipo Pirani, in grado di leggere la pressione totale esistente nel sistema. Il vacuometro sopramenzionato ha un campo di lettura che corrisponde alla gamma di pressioni di lavoro della liofilizzazione.

Gli impianti sono completati da strumenti di lettura e registrazione delle temperature dei prodotti, delle piastre termiche, del condensatore, del fluido di riscaldamento; normalmente si usano termometri a resistenza o a termocoppia.

Adeguati sistemi elettronici consentono la regolazione dell'andamento del ciclo in funzione a parametri guida prefissati facendo intervenire, inoltre, dispositivi di sicurezza in caso di necessità a salvaguardia sia del prodotto sia delle singole macchine.

Le varie fasi operative manuali possono essere completamente automatizzate e programmate nel tempo, allo scopo di ridurre al minimo gli interventi manuali.

Per la corretta progettazione e rispondenza funzionale di un liofilizzatore deve preesistere una esatta conoscenza delle caratteristiche chimico-fisiche e, se del caso, biologiche dei materiali da trattare e di eventuali vincoli a esse collegate.

Conoscendo quanto sopra, si stabilisce il trattamento piú idoneo per i materiali fissando le condizioni ottimali di lavoro che faciliteranno il processo.

Facciamo qualche esempio pratico: la liofilizzazione, come abbiamo avuto modo di accennare, è applicabile a tutti quei preparati, composti o prodotti in cui sia da eliminare la componente liquida, di norma acqua, sia essa sotto forma di solvente o costituente.

Per facilitare la liofilizzazione delle soluzioni è possibile variare il punto finale di congelamento (quanto piú esso è alto tanto piú si facilita la sublimazione); questo si può ottenere, chimicamente, variando il pH della soluzione o aggiungendo sostanze « inerti » dette di « supporto », sottoponendo fisicamente i prodotti a particolari tecniche congelanti in grado di evitare sopraffusioni interne oppure diminuendo semplicemente lo spessore dei prodotti.

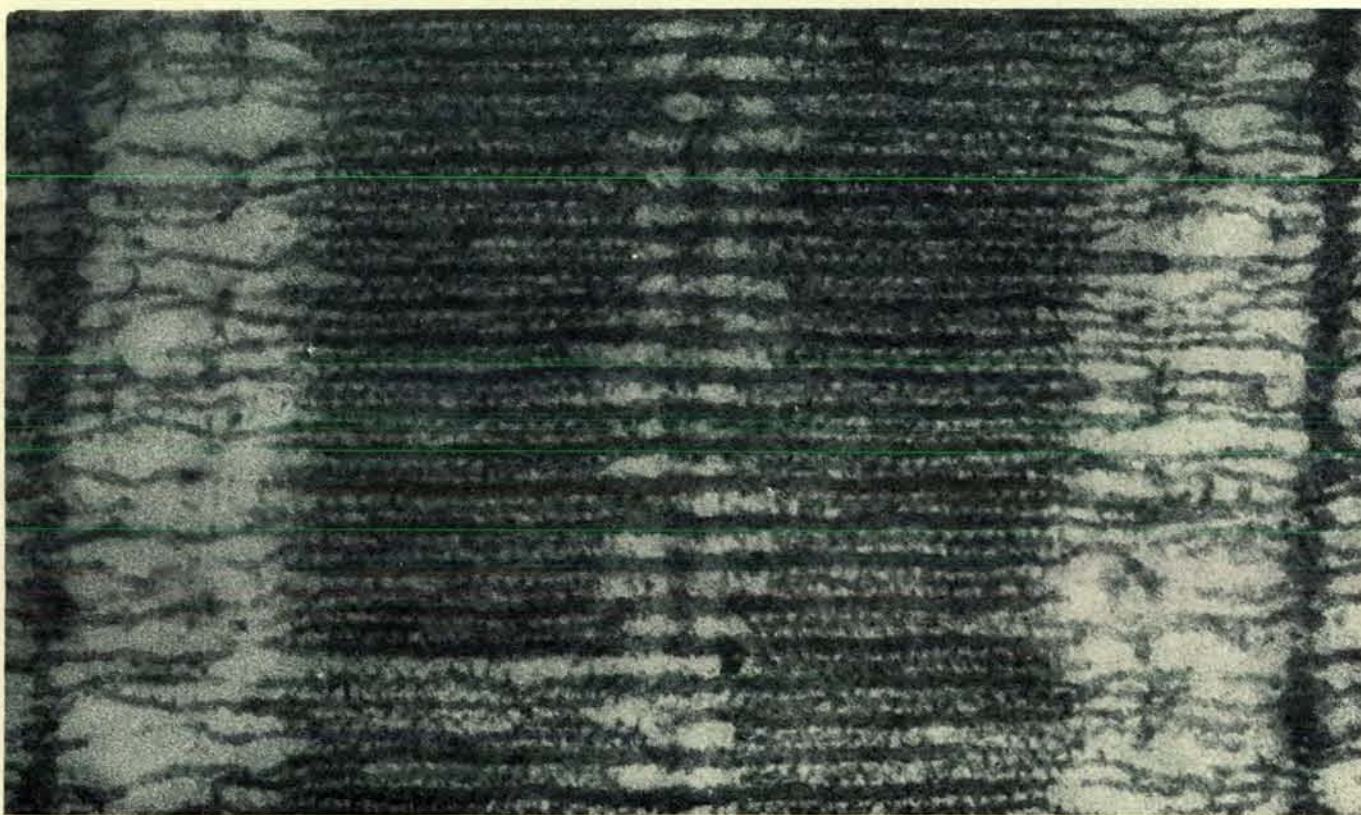
Il costo della liofilizzazione può variare sostanzialmente in dipendenza della produttività ottenibile; tanto piú è alta la resa del prodotto finito tanto piú decresce l'incidenza del costo unitario.

Inoltre l'energia coinvolta in un processo di liofilizzazione è collegata ai quantitativi della fase liquida da eliminare. Per cui, diminuendo la percentuale del solvente da sublimare, diminuiscono le potenzialità energetiche da prevedersi sull'impianto.

Per i prodotti solidi (carni, vegetali, ecc.) ovviamente non si può agire sulla concentrazione o sul punto di congelamento, ma è possibile facilitare la sublimazione operando sulla prima fase del processo: quella del congelamento, o eliminando l'eventuale impermeabilizzazione del fronte esterno della sublimazione con tagli superficiali o pretrattando il materiale.

La metodologia operativa deve consentire la piena armonizzazione tra le esigenze dei prodotti e le prestazioni funzionali degli impianti: particolarmente utile è la razionalizzazione dei metodi di carico, di riscaldamento, e, quando necessario, di integrazione dei cicli di liofilizzazione nella catena produttiva.

La fattiva collaborazione di tecnici altamente specializzati non potrà che ripercuotersi positivamente sullo sviluppo tecnologico della liofilizzazione risolvendo i problemi, specie economici, a essa collegati e parzialmente vincolanti le attuali applicazioni industriali.

Fibra muscolare ottenuta da una sezione longitudinale di muscolo scheletrico di vertebrato ingrandita circa 25 000 volte. La microfotografia è di H. E. Huxley del Medical Research Council di Cambridge, Inghilterra. La fibra è costituita da un insieme di filamenti sottili (*le piú fini strutture orizzontali*) che si estendono nelle due direzioni, da strutture proteiche a forma di disco, le linee Z (*linee spesse verticali*) e da filamenti di maggiori dimensioni posti tra una serie e l'altra di filamenti sottili.

Filamenti di una fibra muscolare ingranditi circa 75 000 volte. Si può osservare che i filamenti grossi sono disseminati di piccole protuberanze, i ponti trasversali. Questi, attaccandosi ai filamenti sottili e ruotando su se stessi, determinano lo spostamento dei filamenti sottili rispetto a quelli spessi, provocando la contrazione muscolare, come dimostra lo schema della pagina a fronte.

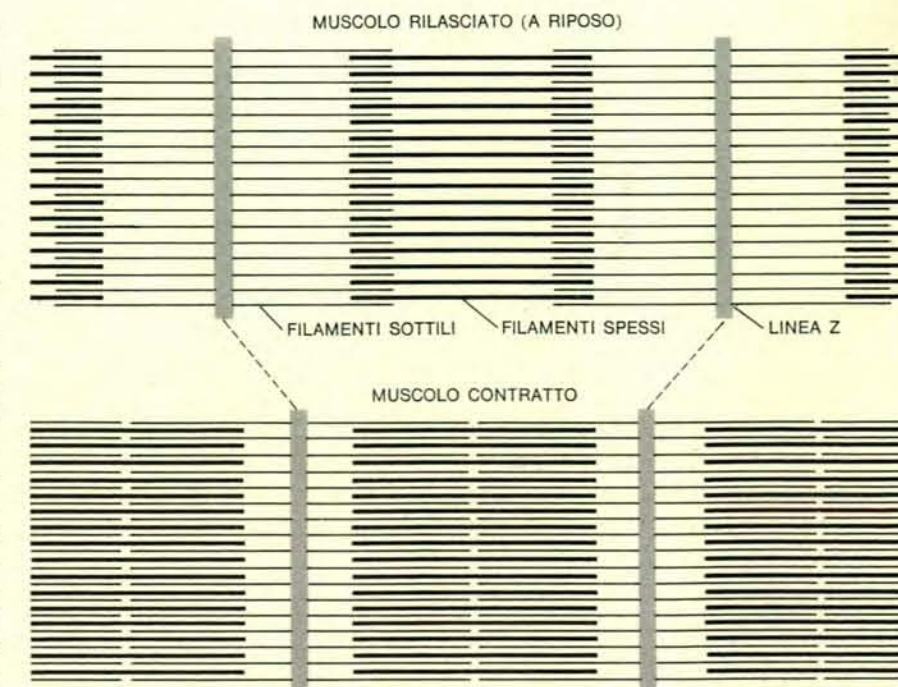
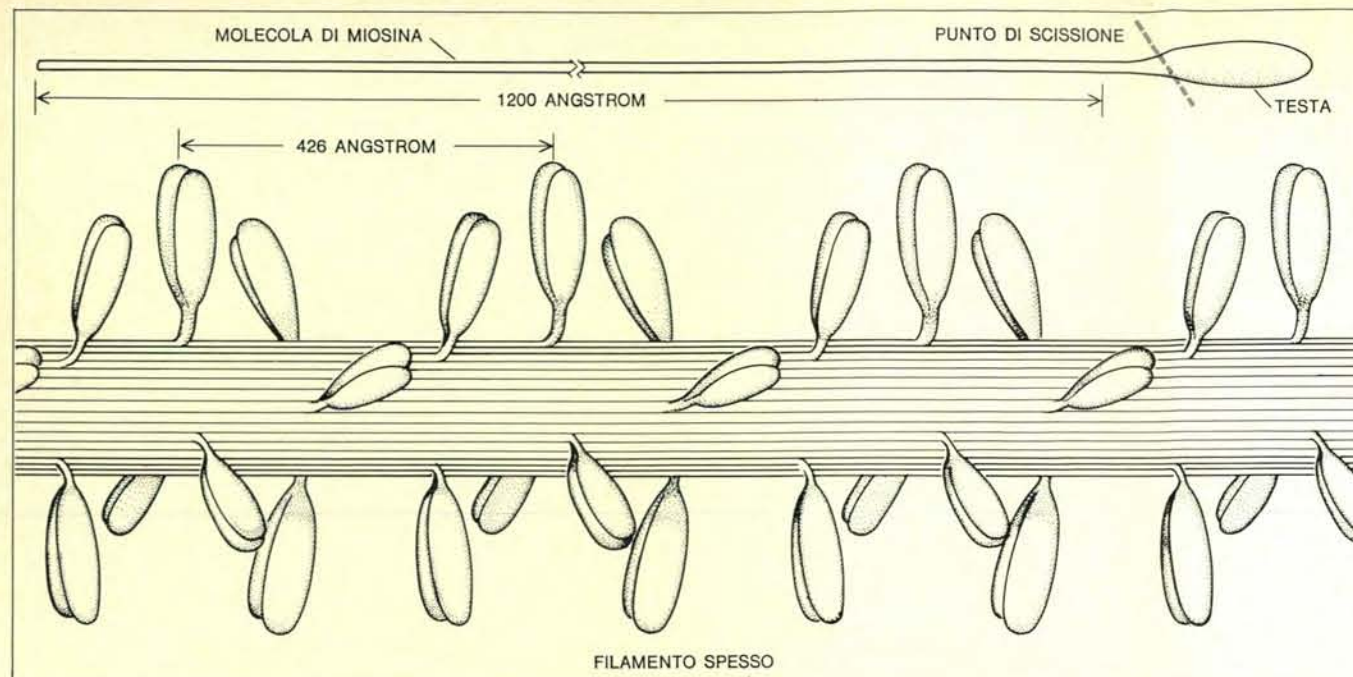
L'azione sinergica delle proteine muscolari

Le fibre muscolari sono costituite da quattro proteine principali: miosina, actina, tropomiosina e troponina. La contrazione del muscolo è funzione della loro interazione che, a sua volta, è controllata dagli ioni calcio

di John Murray e Annemarie Weber

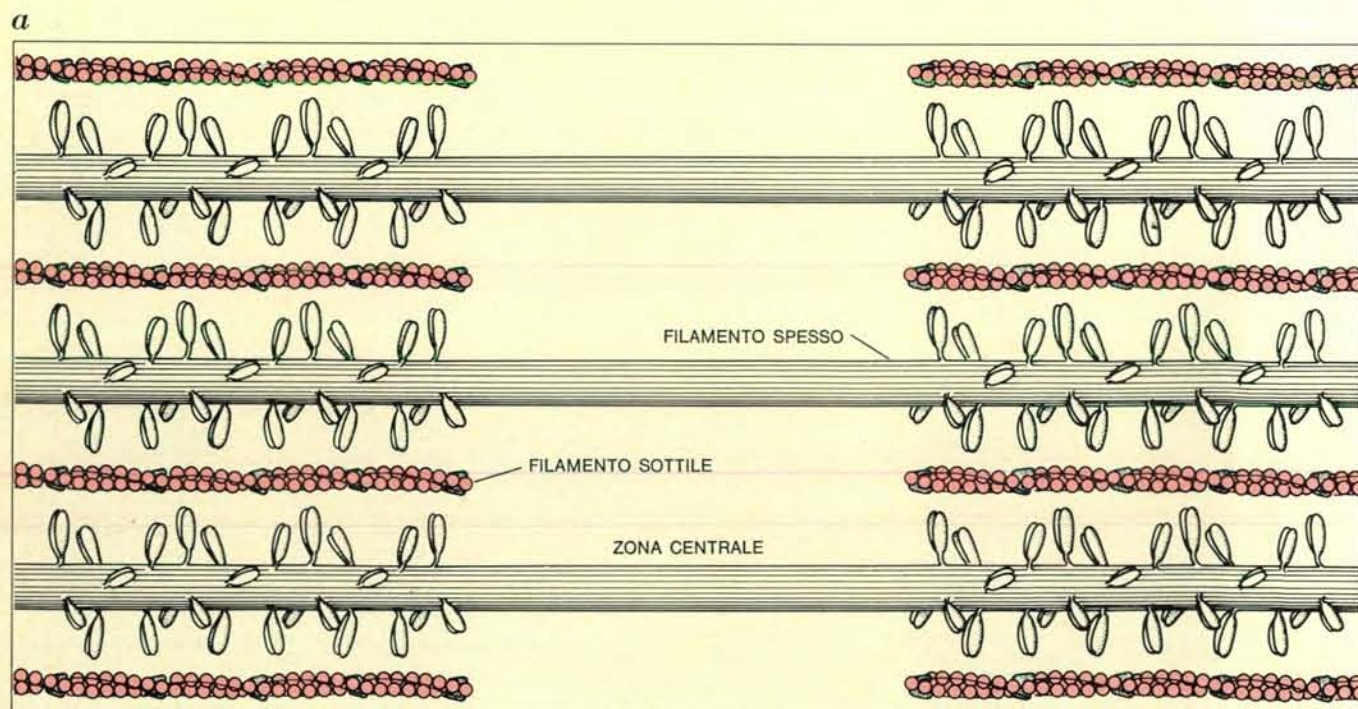
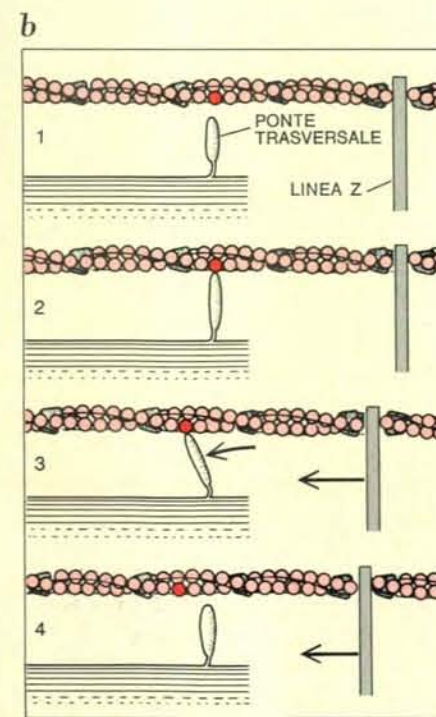
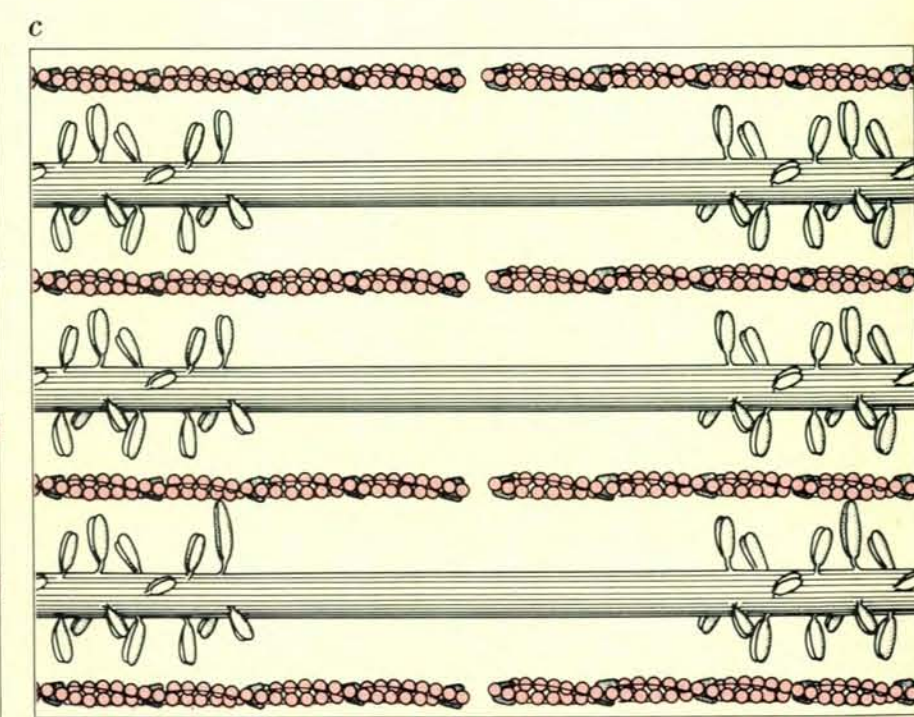
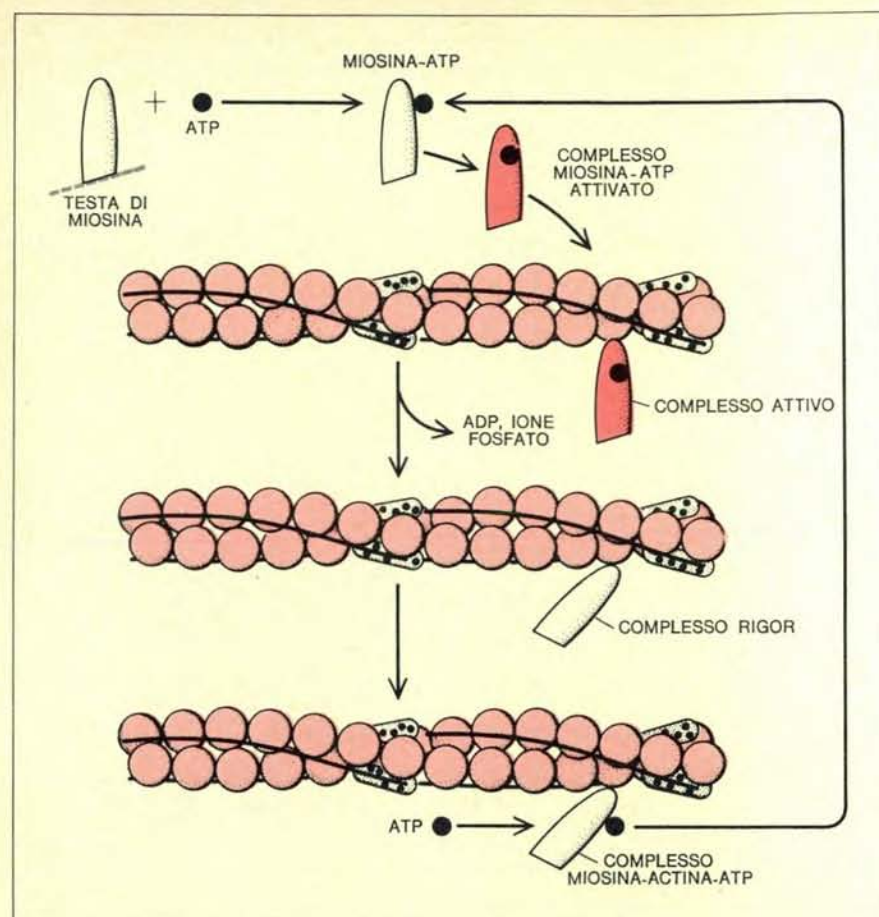
Ogni attività della cellula vivente dipende in larga misura dal comportamento di proteine specifiche. Una dimostrazione evidente di questa importante generalizzazione è fornita dalla contrazione delle fibre muscolari, che può essere spiegata essenzialmente sulla base delle proprietà di quattro proteine: miosina, actina, tropomiosina e troponina. Ognuna di queste proteine può essere ottenuta dal muscolo in forma pura e studiata separatamente, ma il loro comportamento dipende sostanzialmente dalle strutture altamente organizzate che esse formano congiuntamente nella cellula muscolare, e dalle interazioni biochimiche cui queste strutture danno luogo.

La conoscenza delle proteine muscolari e delle strutture che esse formano è dovuta sostanzialmente al lavoro di Albert Szent-Györgyi negli anni 40 e piú recentemente di H. E. Huxley al Laboratorio di biologia molecolare del Medical Research Council a Cambridge in Inghilterra, della scomparsa Jean Hanson al Kings College di Londra, e di Setsuro Ebashi all'Università di Tokyo. Nella cellula muscolare le quattro proteine sono organizzate in due sistemi multimolecolari, i filamenti sottili e grossi, che sono i componenti di base della fibra muscolare. I filamenti possono essere preparati in laboratorio, sia dissociando delicatamente una fibra nei suoi componenti, sia permettendo alle proteine precedentemente purificate, di combinarsi, formando con ciò i filamenti. Un modo di capire il contributo di ciascuna proteina alla struttura finale consiste nel descrivere i singoli componenti molecolari e le loro interconnessioni.

Il piú grosso dei due tipi di filamenti che si osservano nelle fotografie al microscopio elettronico del muscolo contiene tutta la miosina del muscolo e piccoli quantitativi di altre proteine la cui funzione non è chiara e che, da qui in avanti, trascureremo. Una molecola di miosina assomiglia a un sottile bastoncello con due piccole protuberanze globulari alle estremità, dette « teste ». (La miosina può essere « decapitata » per mezzo di un enzima che rompe i legami chimici che tengono unita la testa al resto della molecola. Questo è il principio di un metodo sviluppato da Susan Lowey della Brandeis University per preparare teste isolate. Queste ultime mantengono intatta la attività chimica della miosina, ma sono molto piú maneggevoli dal punto di vista sperimentale.) Nei filamenti piú spessi le molecole di miosina sono sistemate in fasci, dando luogo a una struttura a forma di sigaro che, in tutta la sua lunghezza, è costellata da protuberanze, con l'eccezione di una zona centrale libera. Al microscopio elettronico queste protuberanze appaiono come minuti ponti che sembrano unire i filamenti sottili a quelli spessi. Apparentemente ciascuna protuberanza è costituita dalla testa di una molecola di miosina; nei filamenti spessi le molecole sono orientate con le teste verso le due estremità, per cui la zona centrale è priva di protuberanze. Un filamento spesso di un muscolo normale è

L'aspetto striato della fibra muscolare è dovuto alla disposizione dei filamenti grossi e sottili. Lo schema rappresenta un sarcomero (l'unità del muscolo, tra due linee Z) e parte dei due adiacenti, pressappoco nella stessa scala della microfotografia della pagina a fronte in alto. I ponti trasversali ruotano in senso opposto ai due lati della zona centrale priva di ponti. Ciò determina il movimento relativo dei due ordini di filamenti, cui è dovuto il passaggio dallo stato di rilasciamento a quello di contrazione.

Il filamento spesso è un aggregato di molecole di miosina, lunghi bastoncelli con una doppia testa a un'estremità. La testa ha un sito attivo dove avvengono gli eventi chimici della contrazione muscolare; le teste possono essere separate dal resto della molecola per mezzo di enzimi, ciò che ne permette lo studio. Nei filamenti spessi le molecole di miosina sono organizzate in fasci lunghi circa 1,5 micron con le teste riunite in gruppi di tre. Il filamento sottile è un aggregato di molecole di actina, tropomiosina e troponina. Le molecole di actina sono piccole e di forma globulare, legate a formare una doppia struttura elicoidale. La tropomiosina, una molecola lunga e sottile, forma un lungo filo situato sulla catena di actine, presso ciascun solco della doppia elica. Una molecola di troponina, di forma globulare è situata all'estremità di ciascuna tropomiosina. Quest'ultima copre sette molecole di actina e ci sono da 300 a 400 molecole di actina nel filamento la cui lunghezza è di circa 1 micron.

lungo circa 1,5 micron e contiene parecchie centinaia di molecole di miosina.

Le altre proteine principali connesse con la contrazione muscolare sono contenute nei filamenti sottili. Delle tre, l'actina è presente in quantità maggiore. Le molecole di actina, sferoidali e di piccole dimensioni, sono sistemate nel filamento sottile in modo da formare una doppia fila di palline, attorcigliata su se stessa. Un'importante caratteristica delle molecole di actina è di non essere simmetriche; ciascuna si comporta come se avesse una parte anteriore diversa dalla parte posteriore. Dato che le molecole di actina sono organizzate in strutture filamentose, in modo che la parte anteriore di ogni molecola è a contatto con la parte posteriore della successiva, l'intero filamento viene ad assumere una polarità determinata; la parte anteriore di tutte le molecole di actina guarda in una direzione, la parte posteriore nella direzione opposta. Questa polarità è essenziale per la contrazione muscolare.

Oltre all'actina i filamenti sottili del muscolo contengono le proteine tropomiosina e troponina. Le molecole di tropomiosina sono lunghe e sottili e sono unite fra loro all'estremità, dando così luogo a un sottile filamento situato sulla superficie delle molecole di actina. Ciascuna fila di molecole di actina ha il suo filamento di tropomio-

I filamenti spessi e sottili si sovrappongono in una struttura ordinata a formare una fibra muscolare. Due gruppi di filamenti sottili si dirigono l'uno verso l'altro a partire da due linee Z contigue. Nello spazio fra due gruppi di filamenti sottili, parzialmente sovrapposto a essi, è sistemato un gruppo di filamenti spessi (*a*). Questa organizzazione rende ragione dell'apparenza striata del muscolo, come si vede nella figura delle pagine precedenti. Le teste delle molecole di miosina costituiscono i ponti trasversali che entrano in contatto con i filamenti sottili. Le teste sono orientate in senso opposto rispetto a una zona centrale priva di ponti trasversali. Il muscolo si contrae quando un impulso nervoso dà origine a una serie di eventi che determina l'attacco delle teste di miosina ai filamenti di actina, la loro successiva rotazione e quindi lo stacco delle teste stesse (*b*). In questo modo la sovrapposizione tra filamenti sottili e spessi aumenta causando l'accorciamento del muscolo (*c*).

Schema degli eventi chimici della contrazione muscolare che si verificano nel sistema sperimentale descritto dagli autori. La testa di una molecola di miosina si combina con una molecola di adenosintrifosfato (ATP). Il complesso miosina-ATP, trasformato in una forma intermedia attivata, si lega a una molecola di actina del filamento sottile. Questo complesso viene idrolizzato: l'ATP si scinde in adenosindifosfato (ADP) e fosfato inorganico con liberazione di energia che, nel muscolo intatto, viene utilizzata per la contrazione. Il risultante complesso *rigor* permane finché una nuova molecola di ATP si lega alla testa della miosina; a questo punto il complesso miosina-ATP ricomincia il ciclo.

sina collocato presso il solco tra le due file appaiate. La troponina ha una forma piú o meno globulare ed è situata sulle molecole di tropomiosina, probabilmente non lontano dall'estremità della molecola. Nel filamento sottile una molecola di tropomiosina è distesa sopra sette molecole di actina, con una molecola di troponina per ogni molecola di tropomiosina. I filamenti sottili del muscolo scheletrico in genere sono lunghi 1 micron e contengono da 300 a 400 molecole di actina e da 40 a 60 molecole di tropomiosina. Nel muscolo intatto un'estremità del filamento sottile è attaccata alla linea Z, una struttura proteica piatta. I filamenti sottili, che si dipartono perpendicolarmente dalle due superfici di questa struttura, hanno polarità opposte ai due lati della linea Z. In un muscolo a riposo le linee Z sono distanziate fra loro di circa 2,2 micron e i filamenti sottili sono allineati parallelamente all'asse maggiore della fibra muscolare.

Per una completa descrizione della struttura della fibra muscolare resta da vedere come i filamenti di maggiori dimensioni siano situati nello spazio tra i filamenti sottili che si dipartono dalle linee Z. I due gruppi di filamenti sono organizzati in sistemi paralleli, in parte sovrapposti (*si vedano le figure in basso alle pagine 46 e 47*). Lo spazio tra i due ordini di filamenti è tale da permettere il reciproco scorrimento senza restrizioni. L'accorciamento del muscolo durante la sua contrazione è dovuto appunto a questo fenomeno. La forza che determina lo scorrimento proviene dai ponti trasversali dei filamenti spessi; essi si attaccano al filamento sottile sotto un certo angolo e successivamente ruotano, facendo scorrere i filamenti sottili rispetto a quelli grossi. I ponti trasversali situati ai due lati della zona centrale del filamento grosso ruotano in senso opposto, come se fossero remi azionati da rematori che manovrano in senso opposto alle due estremità di una barca. Il risultato di questo movimento rotatorio è quello di aumentare la sovrapposizione fra i due ordini di filamenti, riducendo cosí la distanza fra le linee Z e accorciando il muscolo.

Per ottenere un accorciamento apprezzabile ciascun ponte trasversale deve agire ciclicamente: deve attaccarsi, ruotare, staccarsi e riattaccarsi in un punto piú distante del filamento sottile. Un singolo ciclo di attacco, rotazione e stacco provoca uno scorrimento del filamento sottile, rispetto a quéllo spesso, di circa 100 angstrom. Ciò significa che, se tutti i ponti trasversali di una fibra muscolare eseguono un solo ciclo, il muscolo si accorcia di circa l'1 per cento rispetto alla sua lunghezza di riposo.

L'energia per la contrazione muscolare, come per la maggior parte delle attività vitali della cellula, proviene dall'idrolisi dell'adenosintrifosfato (ATP). Questa reazione scinde l'ATP, ad alto livello energetico, in composti a minor contenuto energetico: adenosindifosfato (ADP) e fosfato inorganico. La differenza in contenuto energetico tra l'ATP e i suoi prodotti di scissione può essere utilizzata per compiere lavoro utile, come determinare la rotazione di un ponte trasversale, spostando il filamento sottile rispetto a quello spesso e accorciando la fibra muscolare.

Perché il muscolo possa contrarsi la presenza di ATP e delle quattro proteine non è sufficiente. È necessaria la presenza degli ioni calcio che, come in molti altri processi cellulari, costituiscono un fattore di regolazione. Poiché la contrazione muscolare (il ciclico attaccarsi e staccarsi dei ponti trasversali, reso energeticamente possibile dall'idrolisi dell'ATP) non può aver luogo in assenza di calcio, in laboratorio la contrazione muscolare può essere attivata (e disattivata) a volontà semplicemente fornendo (e togliendo) ioni calcio. Nell'organismo vivente la concentrazione muscolare è regolata esattamente in questo modo. Quando l'impulso nervoso che dà il via alla contrazione arriva alla cellula muscolare, gli ioni calcio, si trasferiscono dalle speciali vescicole in cui erano contenuti nel reticolo sarcoplasmico (un sistema intracellulare di tubuli membranosi), nel liquido che circonda i filamenti, dando inizio all'attività ciclica dei ponti trasversali e quindi alla contrazione muscolare. Il calcio viene rapidamente rimosso e reimmesso nelle vescicole di contenimento da una speciale pompa situata sulle membrane del reticolo. La rimozione del calcio blocca l'attività ciclica dei ponti trasversali e il muscolo si rilascia. Sia la fuoriuscita del calcio dalle vescicole, sia la successiva rimozione sono processi molto

rapidi che, normalmente, avvengono in frazioni di secondo.

Numerose tecniche biochimiche sono molto adatte allo studio di soluzioni di proteine, dove le singole molecole sono uniformemente disperse. Tali soluzioni possono essere facilmente preparate da cellule di tessuti come il rene o il fegato. Queste tecniche tuttavia sono meno indicate per lo studio di aggregati ordinati di proteine, come i filamenti di actina e miosina. Benché le proteine muscolari possano essere isolate in forma pura, l'accorciamento del muscolo è dovuto a una speciale organizzazione tridimensionale di queste proteine; quando esse vengono isolate e disperse in soluzione per studiarle in modo piú semplice, le manifestazioni principali della loro attività – l'accorciamento e la generazione di tensione – sono irreversibilmente perdute. In conclusione quindi né il mantenimento completo della organizzazione molecolare, né la completa dispersione delle molecole proteiche consentono uno studio soddisfacente. Il metodo che adottammo con Robert D. Bremel presso le Facoltà di medicina dell'Università di St. Louis e che ora applichiamo all'Università di Pennsylvania è un compromesso. Noi distruggiamo parzialmente la struttura muscolare, separando i filamenti grossi e isolando le teste e il materiale delle linee Z ma lasciando i filamenti sottili intatti, anche se in disordine. In questo modo le proteine sono sufficientemente disperse da poter essere studiate per mezzo delle comuni tecniche applicate alle soluzioni omogenee. Una caratteristica delle proteine muscolari, che le rende adatte a studi dettagliati, consiste nel mantenere le loro proprietà chimiche (la capacità di scindere l'ATP), nonostante la perdita dell'organizzazione strutturale e quindi delle manifestazioni meccaniche (la capacità di accorciarsi e di sviluppare tensione). Questa caratteristica rende possibile lo studio indiretto dell'attività contrattile in varie condizioni sperimentali, considerando la velocità di scissione dell'ATP, invece che misurando direttamente lo sviluppo di tensione o l'accorciamento.

Una delle prime proprietà del muscolo da studiare con questo metodo era l'attività di regolazione esercitata dagli ioni calcio. La scissione dell'ATP in una soluzione di teste di miosina e di filamenti sottili era assolutamente dipendente dalla presenza di ioni calcio, come d'altra parte ci si poteva aspettare dal comportamento del muscolo integro. Ebashi aveva scoperto che l'attività di regolazione del calcio a sua volta dipendeva dalla presenza delle proteine troponina e tropomiosina sul filamento di actina. In presenza di filamenti di actina preparati artificialmente, privi di queste due proteine, la contrazione diviene assolutamente insensibile al calcio, e l'idrolisi dell'ATP continua in maniera incontrollata, fino a completo esaurimento dell'ATP. (Questo comportamento è comune a tutti i muscoli dei vertebrati, ma non è la sola forma di regolazione presente in natura. Andrew G. Szent-Györgyi scoprí che tra gli invertebrati esiste una diversa forma di regolazione da parte degli ioni calcio.) Per capire in qual modo la troponina e la tropomiosina rendono il sistema sensibile agli ioni calcio, è necessario discutere un po' piú dettagliatamente gli eventi chimici della contrazione muscolare.

Le reazioni chimiche che danno luogo all'idrolisi dell'ATP avvengono sulle teste delle molecole di miosina, che costituiscono i ponti trasversali che si dipartono dal filamento spesso. La scissione procede in varie fasi distinte. Dapprima una molecola di ATP si lega in un particolare sito sulla superficie della proteina. La tendenza dell'ATP a legarsi in questa regione è cosí elevata che, di solito, in un muscolo intatto praticamente ogni testa di miosina ha un ATP legato.

Il secondo passo nell'idrolisi dell'ATP è la trasformazione di questo complesso miosina-ATP in una forma intermedia attivata. A differenza del complesso miosina-ATP da cui deriva, questa forma intermedia ha una grande tendenza a legarsi a una molecola di actina del filamento sottile.

Benché la struttura intermedia sia sufficientemente stabile quando è lasciata a se stessa, non appena si lega a una molecola di actina essa si scinde rapidamente, con idrolisi dell'ATP e liberazione di energia. Nel sistema sperimentale in soluzione l'energia liberata dalla scissione dell'ATP è persa in calore. Nel muscolo intatto invece questa energia è utilizzata per far ruotare il ponte trasversale, determinando con ciò lo scorrimento del filamento sottile rispetto a quello grosso e di conseguenza l'accorciamento del muscolo. Lo stadio finale della reazione di idrolisi, lo stacco del ponte trasversale della molecola di actina, avviene solo quando un nuovo ATP si lega al complesso miosina-actina. Il complesso che ne risulta, miosina-actina-ATP si dissocia rapidamente in una molecola di actina libera e in un complesso miosina-ATP «non attivato» *(si veda la figura a pagina 48)*.

La regolazione della contrazione muscolare è dovuta principalmente agli ioni calcio che, probabilmente legandosi alla troponina, determinano il passaggio del filamento sottile da uno stato di riposo «off» a uno di attività «on». Quando il filamento sottile è in stato «off» *(in alto a sinistra)* un complesso intermedio attivato miosina-ATP non può legarsi a esso; quando il filamento è in stato «on» *(in basso a sinistra)* può formarsi un complesso attivo *(in basso a destra)* con susseguente idrolisi. La formazione dei complessi *rigor* tuttavia non è sotto il controllo del calcio. Sia che il filamento sottile sia nello stato «on» o «off», una testa di miosina può combinarsi con una molecola di actina a formare un complesso *rigor* *(in mezzo a destra)*.

Quando la concentrazione di ATP è bassa i complessi *rigor* possono assumere funzioni di controllo. I complessi intermedi attivati miosina-ATP non si combinano col filamento sottile in stato «off», ciò che invece avviene per le teste di miosina *(in alto)*. Se si sono formati molti complessi *rigor* (fenomeno che avviene solo se non c'è molto ATP, che stacca le teste di miosina) il filamento sottile è in stato «on» *(al centro)*, i complessi intermedi attivati si combinano con l'actina, e si ha idrolisi *(in basso)*.

Il lettore si sarà reso conto che, nel corso del processo di idrolisi, si formano due tipi di complessi miosina-actina, uno ad alto livello energetico, l'altro a basso. Il primo si forma quando il complesso intermedio attivato miosina-ATP si lega a una molecola di actina. Questo complesso ad alto livello energetico è di breve durata: in un centesimo di secondo circa, l'ATP si scinde con liberazione di energia e il complesso decade a basso livello energetico. Questo rimane intatto fino all'arrivo di una nuova molecola di ATP. Nel muscolo intatto, dove si ha abbondanza di ATP, ciò avviene in un tempo brevissimo (forse un millesimo di secondo). Quando non c'è ATP, tuttavia, il complesso a basso livello energetico è effettivamente molto stabile, come si può dimostrare in laboratorio dove tali complessi si possono ottenere facendo combinare molecole di actina e miosina in assenza di ATP. In queste condizioni il complesso a basso livello energetico non ha origine dal decadimento di un complesso ad alto livello energetico, ma dalla combinazione diretta di una molecola di actina con una miosina priva di ATP. I complessi cosí formati possono essere mantenuti a lungo e quindi dissociati con l'aggiunta di ATP.

La stabilità dei complessi a basso livello energetico, in assenza di ATP spiega la rigidità cadaverica o *rigor mortis*. La graduale scomparsa di ATP che segue la morte impedisce a un numero sempre maggiore di molecole di miosina di legare una molecola di ATP e quindi conduce alla formazione di un numero sempre maggiore di complessi stabili, a basso livello energetico, tra filamenti sottili e spessi. Poiché c'è molto poco ATP per dissociare i complessi stabili una volta formati, la durata media di essi è molto lunga e il muscolo diventa sempre meno estensibile. Per questo motivo i complessi a basso livello energetico sono denominati complessi *rigor*. Al contrario la combinazione dell'actina con il composto intermedio attivato miosina-ATP è denominata complesso attivo per indicare il suo ruolo di generatore di forza.

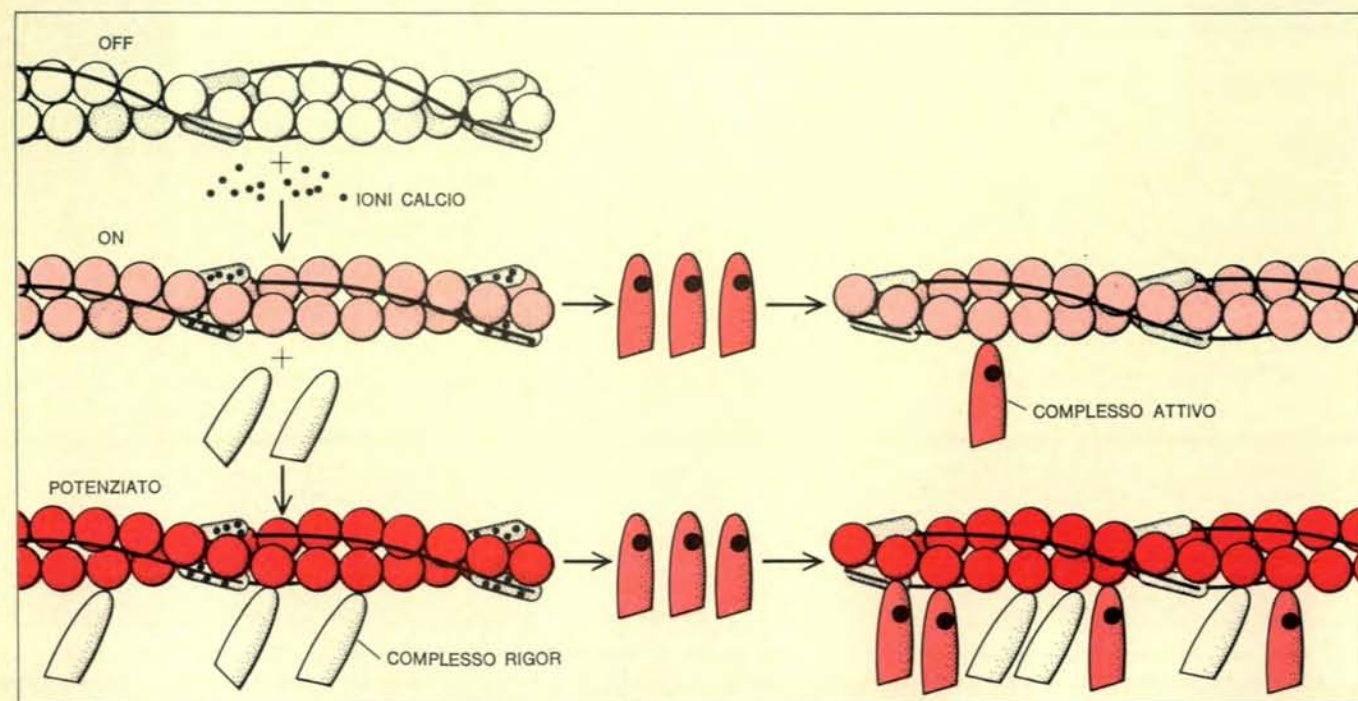
Come avviene che tutta questa serie di reazioni è regolata dagli ioni calcio, la cui assenza provoca un blocco completo di ogni attività? Dato che i vari stadi sono collegati ciclicamente, il blocco di uno qualsiasi di essi inibisce l'intero processo. I dati sperimentali di numerosi laboratori hanno dimostrato che lo stadio sensibile al calcio è la combinazione del complesso intermedio attivato miosina-ATP con l'actina, ossia la formazione dei complessi at-

tivi. Dato che la troponina e la tropomiosina sono indispensabili per l'attività di regolazione del calcio, il blocco della formazione di complessi attivi in assenza di calcio è verosimilmente mediato da queste due proteine. Alla luce delle relative proporzioni delle tre proteine che costituiscono il filamento sottile, ci si può chiedere come una molecola di tropomiosina e una di troponina possano controllare l'attività di sette molecole di actina.

La localizzazione della molecola di tropomiosina nel filamento suggerisce una possibile spiegazione. È noto che il calcio si lega alla troponina; se la risposta di questa proteina alla perdita del calcio a essa legato (che avviene quando l'attività della pompa del calcio rimuove gli ioni dal fluido che circonda i filamenti) è un movimento tale da allontanare la tropomiosina dal solco tra le due file adiacenti di molecole di actina, allora l'effetto di questo spostamento potrebbe essere quello di bloccare fisicamente l'avvicinamento del complesso intermedio attivato miosina-ATP. Questo modello semplice e attraente, proposto recentemente da H. Huxley e J. Haselgrove, sulla base di studi mediante la diffrazione dei raggi X, va modificato leggermente per tener conto di un'ulteriore osservazione: a differenza dei complessi attivi, la formazione di complessi *rigor* è insensibile al calcio, cosí che, quando la miosina e l'actina vengono mescolate in assenza di calcio, i complessi *rigor* si formano ugualmente. Benché il filamento sottile debba essere in uno speciale stato che chiamiamo « on » (acceso), in genere innescato dal legame del calcio con la troponina, per formare il complesso attivo, i complessi *rigor* possono formarsi sia in questo stato che in quello opposto che chiamiamo « off » (spento); la miosina e il complesso intermedio attivato miosina-ATP sono specie chimicamente distinte (*si veda la figura a pagina 50*).

Apparentemente le molecole di actina sono in grado di distinguere una molecola di miosina priva di ATP (che reagisce dando luogo al complesso *rigor*) da un complesso intermedio attivato (che reagisce dando luogo a un complesso attivo). Concludemmo che in condizioni sperimentali in cui i filamenti sottili si trovino in presenza di ambedue le forme di miosina (come succede quando la concentrazione di ATP è bassa rispetto alla concentrazione di miosina), in assenza di calcio dovrebbero formarsi solo complessi *rigor* e quindi non ci dovrebbe essere idrolisi di ATP. Ripetemmo questo esperimento parecchie volte in modi diversi, sempre con lo stesso risultato: contrariamente alle previsioni, l'idrolisi dell'ATP aveva luogo perfettamente! Questa sconcertante osservazione fu solo la prima di molte altre che indicavano chiaramente che avevamo interpretato il comportamento dei filamenti sottili durante la contrazione muscolare in modo troppo semplicistico.

I fenomeni sopra descritti dimostrano una perdita della capacità di regolazione. Dato che il calcio era assente, il filamento sottile avrebbe dovuto essere in condizioni « off » rispetto all'interazione con il complesso attivato miosina-ATP, tuttavia l'osservata idrolisi dell'ATP indicava inequivocabilmente che le due proteine interagivano. Una prima traccia per spiegare questa apparente perdita di controllo ci fu data da un esperimento in cui paragonammo il grado di perdita della regolazione a differenti livelli di concentrazione di ATP. Trovammo che, a concentrazioni di ATP elevate, la regolazione dell'attività contrattile da parte del calcio era normale; abbassando il livello di ATP la perdita della regolazione diventava sempre piú grave, finché, a concentrazione di ATP molto basse, il calcio perdeva completamente ogni capacità di controllo.

A cosa può essere dovuto questo comportamento? Un effetto diretto dell'ATP sulla capacità del calcio di legarsi non sembrava probabile in quanto l'ATP non interagisce affatto con le due proteine regolatrici, troponina e tropomiosina. Fummo quindi obbligati a cercare la spiegazione in qualche effetto indiretto dell'ATP. Dato che il legame dell'ATP con un complesso *rigor* provoca la sua rapida dissociazione in miosina-ATP e actina, il numero di complessi *rigor* presenti sul filamento sottile in ogni istante deve essere tanto piú basso, quanto piú è al-

Lo stato potenziato è una terza possibile condizione per l'actina. Le molecole nello stato « off » sono attivate dagli ioni calcio e possono quindi formare complessi attivi (*a destra*). A basse concentrazioni di ATP si formano complessi *rigor*. Questo porta le molecole di actina nello stato « on » a uno stato potenziato in cui viene legato un numero maggiore di complessi intermedi attivati.

ta la concentrazione di ATP. Esisteva quindi la possibilità che la formazione di complessi *rigor* ponesse in qualche modo il filamento sottile in condizioni « on »; ossia permettesse la formazione di complessi attivi, e quindi l'idrolisi dell'ATP, anche in assenza di calcio.

Questa possibilità richiedeva una nuova serie di esperimenti. Infatti, se la perdita della regolazione del calcio a basse concentrazioni di ATP era dovuta solo al gran numero di complessi *rigor* sul filamento sottile e non a qualche altro effetto sconosciuto dell'ATP, allora una variazione del numero di questi complessi, a concentrazione di ATP costante, dovrebbe far variare il grado di indipendenza dalla calcio-regolazione. Il numero di complessi *rigor* sul filamento sottile può essere variato semplicemente cambiando il rapporto di concentrazione actina-miosina, esperimento che può essere fatto a concentrazione di ATP costante. Quando eseguimmo questa prova trovammo che la regolazione mediante il calcio sparisce ad alti rapporti miosina-actina, e ricompare a bassi rapporti miosina-actina, esattamente come ci si può attendere sulla base del presupposto che i complessi *rigor* pongano il filamento in condizione « on » anche in assenza di calcio.

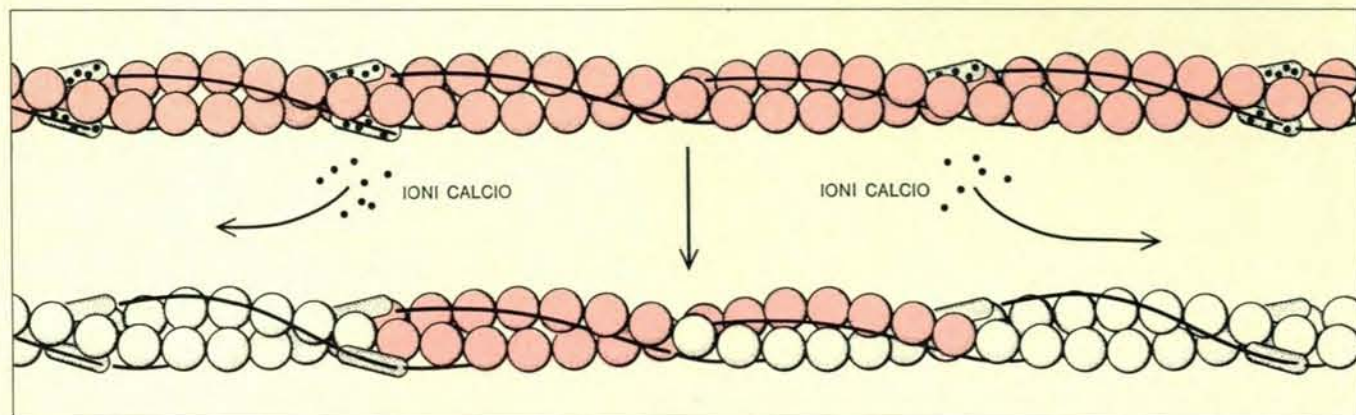
Mentre era confortante vedere la predizione confermata, ciò comportava parecchie inaspettate conseguenze. Il controllo del filamento sottile non poteva piú essere considerato come un semplice processo « on-off », mediato solo dalla troponina e sotto il controllo dei soli ioni calcio. Questa conclusione fu dimostrata anche in altri esperimenti. Un'ulteriore conclusione fu che il livello di attività dell'actina è variabile. Infatti quando gli effetti dei complessi *rigor* furono studiati in dettaglio noi trovammo che essi, non solo ponevano in stato « on » le molecole di actina prima « off », (in assenza di calcio), ma anche, in presenza di calcio, modificavano il comportamento delle molecole in actina che già si trovavano in stato « on ». Questa modificazione consisteva in una velocità di idrolisi dell'ATP molto piú elevata che di norma; ciò che nel nostro sistema è indice di una piú intensa attività contrattile. Noi chiamiamo potenziato questo stato dell'actina, per distinguerlo dallo stato « on » in cui le molecole di actina sono sí attive (in quanto il calcio è legato alla troponina), ma manifestano un'attività piú bassa, in quanto non sono in presenza di complessi *rigor*. È stato dimostrato che il livello piú alto di attività contrattile nello stato potenziato è dovuto a una elevata velocità di formazione di complessi attivi a partire dall'actina e dal complesso attivato miosina-ATP (*si veda la figura in questa pagina*).

Questi e altri esperimenti ci spinsero a riconsiderare le nostre idee sul comportamento delle molecole di actina nel filamento sottile. Le numerose file di molecole di actina che costituiscono un filamento sottile erano state considerate come entità piú o meno indipendenti, riunite in una struttura unica semplicemente per fornire un conveniente sostegno ai ponti trasversali dei filamenti spessi su cui far scorrere i filamenti sottili e quindi accorciare il muscolo. La possibilità di accedere a questi sostegni era sotto il controllo di una molecola di troponina e una di tropomiosina, per ogni sette di actina. Le molecole di actina vicine tra loro sullo stesso filamento erano considerate praticamente indipendenti, a parte il fatto di essere sotto il controllo delle stesse molecole di troponina, né si pensava che avessero alcuna influenza sulle reciproche attività. Questi concetti evidentemente non reggevano piú. Se i complessi *rigor* potevano influenzare l'attività del filamento sottile, doveva di necessità esserci comunicazione tra le molecole di actina. Ciò in quanto l'actina direttamente legata in un complesso *rigor* non può esercitare alcuna attività idrolitica (una molecola di actina può legarsi solo con una testa di miosina alla volta, sia essa attivata o no); l'informazione circa la presenza di un complesso *rigor* deve essere trasmessa a una molecola di actina libera di esercitare attività idrolitica.

Il nuovo concetto di comunicazione fra molecole di actina lungo il filamento sottile si dimostrò molto utile nel suggerire una nuova serie di problemi (o almeno un modo nuovo di formulare problemi vecchi). Come si trasmette l'informazione lungo il filamento? A quale distanza massima può venire trasmesso un messaggio; da un'estremità all'altra, o solo a gruppi di molecole di actina controllate dalla stessa molecola di troponina? Le informazioni trasmesse riguardano unicamente il numero dei complessi *rigor*; o è possibile trasmettere altri messaggi? Considerando quest'ultimo problema ci venne l'idea che la regolazione attraverso il calcio mediata dal legame con la troponina non fosse che uno degli aspetti della regolazione mediata dal numero dei complessi *rigor*. Forse la troponina trasmetteva alle sette actine sotto il suo controllo il messaggio: « calcio legato, entrare in attività », esattamente nella stessa maniera con cui il messaggio circa la formazione di complessi *rigor* è trasmesso da un'actina all'altra. Se le cose stanno realmente cosí, dovrebbe esistere la possibilità di trasmettere messaggi anche nel verso opposto: dall'actina alla troponina. Studiando questo problema trovammo infatti che un tal genere di informazioni veniva in effetti trasmesso alla troponina; la troponina infatti veniva informata della formazione di complessi *rigor* e rispondeva modificando sostanzialmente il suo comportamento: ossia legando piú calcio. A questo punto era evidente che piú di un tipo di informazione poteva essere trasmesso lungo il filamento sottile; quindi cercammo e trovammo le prove di altri tipi di informazioni. Risultò evidente che le molecole di actina del filamento sottile erano informate, in ogni istante, circa il numero dei complessi attivi presenti; la risposta a questa informazione era analoga a quella evocata, sul filamento attivato dal calcio, dalla presenza di complessi *rigor*: veniva potenziata l'attività dello stato « on ».

Avendo stabilito che diversi tipi di informazioni possono essere trasmessi lungo il filamento sottile, passammo a considerare come ciò potesse avvenire. La prima ipotesi era che le molecole di actina trasmettessero informazioni direttamente alle loro immediate vicine. Questa ipotesi poteva venire verificata dato che è possibile preparare filamenti che contengono solo actina, eliminando completamente troponina e tropomiosina; filamenti di tal genere sono stati studiati da Evan Eisenberg al National Heart and Lung Institute e da Carl Moos all'Università di stato di New York a Stony Brook. Data la carenza di troponina questi filamenti non sono piú sotto il controllo del calcio, ma sono sempre in attività. In questi filamenti non è quindi possibile studiare il tipo di informazione che determina il passaggio allo stato « on », è invece possibile studiare il tipo di informazione che determina il potenziamento dell'attività. Questo era del tutto assente. Né complessi *rigor*, né complessi attivi avevano alcun effetto sulle molecole di actina adiacenti; ciascuna molecola era completamente indipendente dal resto del filamento. Quando aggiungemmo la tropomiosina ai filamenti che ne erano privi (senza aggiungere la troponina) ricomparve la risposta di potenziamento. Era quindi chiaro che la tropomiosina ha un importante ruolo nella trasmissione delle informazioni.

La troponina trasmette direttamente il messaggio, o agisce semplicemente modificando le molecole di actina cosí che in sua presenza possono comunicare, mentre sono isolate in sua assenza? Non siamo ancora in grado di dare una risposta esauriente a questa domanda, ma abbiamo almeno qualche indicazione circa una possibile soluzione. Preparammo filamenti sottili con il normale contenuto di tropomiosina, ma senza troponina. Aggiungemmo quindi troponina a dosi crescenti misurando contemporaneamente l'estendersi della regolazione mediante il calcio lungo il filamento di actina. Il motivo di questo esperimento era il seguente: se l'informazione circa il legarsi del calcio è trasmessa direttamente dall'actina, con la tropomiosina come semplice catalizzatore, era lecito aspettarsi che l'informazione venisse trasmessa lungo una distanza maggiore di quella di una molecola di tropomiosina legata con una di troponina. Al contrario se la tropomiosina stessa trasmetteva l'informazione direttamente, era logico attendersi che l'informazione venisse recepita solo dalle actine rivestite da una molecola di tropomiosina su cui si fosse collocato una molecola di troponina. Quest'ultima previsione fu confermata. I risultati delle misure dell'attività idrolitica, in relazione all'aggiunta di troponina, dimostrarono che la calcio-regolazione esisteva solo in quelle molecole di actina direttamente collegate sia con una molecola di tropomiosina, sia con una di troponina. Le altre molecole di actina, anche se associate con la tropomiosina, non rispondevano alla troponina presente a distanza. La spiegazione piú semplice in questo caso è che le molecole di actina non comunicano direttamente, ma solo indirettamente tramite la tropomiosina. Questa spiegazione è in accordo con l'osservazione di Haselgrove e Huxley che, quando il filamento di actina è attivato dal calcio o da complessi *rigor*, la tropomiosina rientra nel solco fra le file di actina globulare, rimuovendo quindi ogni ostacolo fra le molecole di actina e le teste della miosina.

Tutti sistemi contrattili studiati fino a oggi contengono tropomiosina, anche quelli in cui manca la troponina. Ciò è piuttosto strano, in quanto la tropomiosina da sola non sembra in grado di influenzare l'attività del filamento sottile. Il ruolo centrale della tropomiosina nei fenomeni descritti suggerisce l'ipotesi che questi fenomeni siano parte integrante del processo contrattile. Forse uno dei motivi per cui la tropomiosina è invariabilmente associata ai sistemi contrattili è la sua capacità di trasformare le singole molecole di actina del filamento sottile in una unità funzionale altamente organizzata.

L'informazione circa la presenza del calcio è trasportata dalla troponina, cui esso si lega, per mezzo della tropomiosina, alle molecole di actina sottostanti la tropomiosina stessa. Ciò fu stabilito con un esperimento in cui si era ridotto il numero di molecole di troponina. Inizialmente il calcio era presente in soluzione (*in alto*). Le molecole di actina sottostanti le molecole di tropomiosina legata con troponina erano quindi in stato « on ». Anche le molecole di actina al di fuori dell'influenza della troponina erano in stato « on » in quanto l'actina, in assenza di troponina, non era piú sotto il controllo del calcio. Quando questo fu eliminato (*in basso*), le sole molecole di actina allo stato « off » furono quelle sottostanti le molecole di tropomiosina.

La natura delle comete

Sebbene la cometa Kohoutek abbia deluso gli osservatori visuali, essa ha fornito una buona occasione per studiare uno di quegli oggetti che sono probabilmente i resti della nube da cui si sono formati il Sole e i pianeti

di Fred L. Whipple

La cometa Kohoutek, come altre comete, è una fontana celeste che zampilla da una grande palla di neve sporca galleggiante nello spazio. La fontana è resa attiva e illuminata dal Sole ed è intensificata enormemente perché sgorga nel vuoto ed essenzialmente in assenza di gravità. Vediamo la fontana come testa e coda della cometa. La coda può estendersi per decine di milioni di chilometri, ma non vediamo mai la palla di neve il cui diametro è solo di pochi chilometri.

La parola « cometa » viene dal greco *aster kometes*, cioè stella dalla lunga capigliatura. La coda della cometa è naturalmente la capigliatura; si può considerare la testa o chioma della cometa come la stella. Entro la chioma c'è la palla di neve: un nucleo di ghiaccio che si muove su una enorme orbita sotto il controllo gravitazionale del Sole. Il nucleo passa quasi tutta la sua vita a grande distanza dal Sole, ibernato nel gelo profondo dello spazio. Quando la sua orbita lo fa ruotare verso il Sole, la sua superficie comincia a sublimare e il gas sublimato si sparge nello spazio. Opponendosi alla debole gravità del nucleo relativamente piccolo, le molecole e gli atomi che emergono trasportano con sé particelle solide. Così il nucleo dà origine alla nube gassosa e polverosa della chioma.

Il Sole illumina la polvere e il gas della chioma, rendendo visibile la cometa. Alcune comete sono molto polverose. La maggior parte della luce che si osserva è semplicemente luce solare diffusa dalla polvere ed è leggermente rossastra. Altre comete contengono poca polvere. Siccome le molecole e gli atomi di un gas diffondono poca luce, queste comete gassose diventano brillanti solo attraverso un doppio processo. Per prima cosa la radiazione ultravioletta del Sole divide le molecole, l'acqua per esempio si dissocia in idrogeno (H) e nel radicale ossidrile (OH). Allora l'atomo o la molecola spezzata possono dare fluorescenza, cioè assorbire la luce solare a una lunghezza d'onda e irradiarla di nuovo alla stessa lunghezza d'onda o (più spesso) a una serie di lunghezze d'onda maggiori.

Quasi tutta la luce delle comete gassose proviene da tali bande di lunghezze d'onda che sono per lo più emesse da molecole spezzate di carbonio, azoto, ossigeno e idrogeno come CH, NH, NH_2, CN e OH e anche C_2 e C_3. Quali sono le molecole originali che, rompendosi, producono questi radicali instabili? Le prime a essere sospettate sono quelle dell'ammoniaca (NH_3) e del metano (CH_4), ma il sospetto non è ancora stato confermato. Ancora più incerta è la situazione riguardo alla molecola che genera il CN. Potrebbe essere il gas cianogeno (C_2N_2)? O il cianuro di idrogeno (HCN)? O forse qualche molecola ancora più strana?

Code di polveri e code di ioni

Indipendentemente dalle risposte a queste domande, si è ora capito che la chioma di una cometa brilla a causa della luce solare diffusa dalla polvere o a causa della luce solare irraggiata da gas fluorescente; di solito per entrambe le cause. La coda di una cometa si crea per un altro effetto del Sole. Le code, come le teste delle comete, hanno una componente gassosa e una polverosa. Per le code polverose l'azione solare non è complicata: la pressione di radiazione della luce solare espelle le particelle di polvere della chioma. Seguendo le leggi del moto dei corpi orbitanti, le particelle di polvere restano indietro rispetto alla chioma via via che si allontanano da essa e formano perciò una coda curva che può essere ricca di particolari.

Molte comete, soprattutto le più brillanti, mostrano una enorme coda che è solo leggermente incurvata. Come il gas della chioma anche queste code brillano per fluorescenza; però le molecole responsabili della radiazione sono ionizzate, cioè sono stati strappati loro degli elettroni e sono rimaste cariche positivamente. In queste code di ioni si trovano monossido di carbonio ionizzato (CO^+), biossido di carbonio (CO_2^+), azoto (N_2^+) e i radicali OH^+ e CH^+, ma non molecole o radicali neutri. La luce solare può ionizzare alcune molecole, ma qual è la pressione capace di respingerle nello

La cometa Kohoutek fu fotografata l'11 gennaio col telescopio Schmidt di 42 centimetri dell'osservatorio Catalina sul monte Lemmon in Arizona. La fotografia è stata fornita da R. B. Minton del Lunar and Planetary Laboratory dell'Università dell'Arizona, che ha fatto numerose riprese della cometa Kohoutek con questo strumento. Questa fotografia è stata scattata da Stephen Kutoroff con una posa di 10 minuti tra 1 ora e 59 minuti e 2 ore e 09 minuti (tempo universale). La linea diagonale che attraversa la fotografia in basso a sinistra è stata lasciata dal passaggio di un satellite artificiale; il crescere e il calare dello splendore della traccia è dovuto al fatto che il satellite ha una forma irregolare e riflette più o meno la luce solare negli ondeggiamenti sulla sua orbita. La fotografia è stata fatta dopo il perielio, punto di massimo avvicinamento al Sole, per cui la cometa si muove praticamente nella stessa direzione in cui la coda della cometa si allontana dalla sua testa. Le osservazioni della cometa alle lunghezze d'onda radio mostrano che essa contiene cianuro di metile (CH_3CN) e cianuro di idrogeno (HCN). Poiché le molecole di questo tipo si formano nelle nubi interstellari, la loro presenza convalida l'ipotesi che le comete abbiano origine in queste zone.

spazio con forze talvolta superiori a 1000 volte la gravità del Sole?

La questione di come si sono formate le code di ioni rimase a lungo un mistero ed è stata risolta solo nell'era delle esplorazioni spaziali. Le sonde spaziali hanno rinviato dati che mostrano come il Sole espella continuamente un milione di tonnellate di gas al secondo, gas che si muove con una velocità radiale di 400 chilometri al secondo. Questo vento solare, che ha una temperatura di un milione di gradi, trascina con sé campi magnetici caotici. I campi sono trasportati da correnti di elettroni nel gas che è quasi completamente ionizzato. Un decennio prima della prima sonda spaziale, Ludwig F. Biermann del Max Planck Institut für Physik di Gottinga dimostrò la necessità di ammettere l'esistenza di qualcosa di simile al vento solare per le code di ioni delle comete. Sebbene la teoria del vento solare delle code di ioni non sia ancora molto precisa, essa indica tuttavia che due processi accoppiano il vento solare al gas cometario.

In primo luogo gli elettroni di alta energia nel vento solare ionizzano le molecole della chioma (insieme alla radiazione solare). In secondo luogo il vento solare dà origine a un'onda arcuata intorno alla chioma. I campi magnetici caotici agiscono ora come rastrelli magnetici che allontanano gli ioni dalla chioma, senza aver effetto sulle molecole e sugli atomi non ionizzati. La forza del vento solare sugli ioni può accelerarli a velocità di parecchie decine di chilometri al secondo in modo che si possono osservare variazioni in una coda di ioni a distanza di molti milioni di chilometri nel breve volgere di mezz'ora.

John C. Brandt del Goddard Flight Center della NASA ha dato una spiegazione per la bella curvatura di queste grandi code. È il risultato del moto trasversale della cometa a varie decine di chilometri al secondo attraverso il moto del vento solare che soffia radialmente dal Sole. La coda di ioni interagisce col vento solare ad alta velocità nello stesso modo in cui il fumo che proviene da un fumaiolo interagisce con l'aria in movimento dando un bell'arco fluttuante sulla Terra.

Residui di comete

Le comete lasciano dietro di sé dei residui nello spazio interplanetario. Parte di essi si vedono dalla Terra come luce zodiacale che è visibile come un chiarore nel cielo a est prima dell'alba e a ovest dopo il tramonto (raggiunge il massimo splendore ai tropici). Gran parte della luce zodiacale vicino al piano dell'orbita terrestre è luce solare diffusa dalla fine polvere che si lasciano dietro le comete. In condizioni ideali di osservazione, la polvere cometaria appare anche come *Gegenschein* o controluce zodiacale: debole macchia luminosa nel cielo notturno in direzione opposta a quella del Sole. Bisogna che le comete contribuiscano con circa 10 tonnellate di polvere al secondo nel sistema solare interno per mantenere questo livello di illuminazione. In un periodo di qualche migliaio di anni le particelle si rompono per effetto di collisioni con altre particelle o sono soffiate via dalla radiazione solare.

Oltre a questo materiale fine, le comete distribuiscono lungo la loro orbita particelle solide piú grandi e noi osserviamo questi meteoroidi come pioggia di meteore quando la Terra passa vicino all'orbita di una cometa regolare. I meteoroidi entrano nella parte superiore dell'atmosfera a una velocità di circa 60 chilometri al secondo e l'attrito atmosferico libera l'energia cinetica dell'oggetto in un bagliore di luce di breve durata. L'energia liberata per grammo di peso del meteoroide supera di molto l'efficienza energetica dei piú potenti esplosivi prodotti dall'uomo. Cosí un oggetto delle dimensioni di un pisello può creare una notevole traccia meteorica.

Non si conosce però nessun meteoroide, associato con una orbita cometaria, che abbia raggiunto il suolo, per cui è realistico domandarsi se le comete abbiano contribuito o meno alle meteoriti raccolte nei nostri musei. Infatti gli studi di meteore con orbite cometarie ci dicono che i detriti cometari sono estremamente fragili. Anche se gli spettri di questi oggetti indicano che i loro principali costituenti sono ferro, magnesio, silicio e altri elementi terrestri, il materiale è tenuto insieme in modo assai debole. I meteoroidi hanno una densità media di meno della metà della densità dell'acqua e si può facilmente romperli con le mani. Entrando nell'atmosfera ad alta velocità il materiale si spezza e sublima. Solo raramente, ammesso che ciò avvenga, giunge sulla Terra in pezzi piú grandi di fine polvere. Si può però raccogliere parte della polvere quando scende lentamente attraverso l'atmosfera.

Banco di ghiaia contro palla di neve sporca

Come si può essere sicuri che il nucleo invisibile di una cometa è veramente una palla di neve sporca? Fino al 1950, quando proposi il modello del nucleo di ghiaccio, l'idea accettata generalmente era quella che il nucleo cometario fosse qualcosa di simile a un banco di ghiaia che viaggiava nello spazio. Si ammetteva che i gas cometari fossero stati assorbiti dalle particelle solide e fossero liberati dalla energia solare quando la cometa si avvicinava al Sole. Siccome si sapeva che le comete contribuiscono alle particelle solide dello spazio interplanetario fino almeno alle dimensioni di un acino d'uva, sembrava ammissibile che queste particelle fossero i principali componenti di una cometa e che i gas fossero di importanza secondaria.

La teoria del banco di ghiaia del nucleo cometario si dimostrò però completamente inadeguata a spiegare tre critiche prove sperimentali. Ciascuna di queste prove è di per se stessa fatale alla teoria del banco di ghiaia, mentre è soddisfatta dall'ipotesi del nucleo di ghiaccio. Le tre prove ci dicono tanto sulla natura delle comete da meritare qualche ulteriore sviluppo.

La prima prova riguarda il fatto che le comete che si avvicinano notevolmente al Sole non vengono distrutte. Nel 1965 si è presentata l'occasione di osservare la cometa Ikeya-Seki che è passata a una distanza dalla superficie del Sole pari a solo un terzo del diametro solare. A tale distanza il calore solare è cosí intenso che dovrebbe fondere e sublimare praticamente tutte le sostanze normali e vi è stato un tempo piú che sufficiente perché qualsiasi oggetto terrestre del diametro di 30 o

La cometa Halley fu fotografata nel maggio 1910 dalla stazione australe del Lick Observatory a Cerro San Cristóbal nel Cile. Siccome ha un periodo orbitale di 76 anni, ritornerà nel 1986. La coda di questa cometa ha una quantità minima di polvere.

La cometa Arend-Roland fu una delle due comete brillanti osservate nel 1957; l'altra cometa era la cometa Mrkos (*si veda la figura nella pagina a fronte*). La fotografia è fatta col telescopio Schmidt di 120 centimetri sul monte Palomar il 30 aprile.

La cometa Mrkos, visibile nell'agosto 1957, indica chiaramente che una cometa ha due tipi di coda come pure che la forma di ciascuna coda cambia col passare del tempo. La coda dritta verso sinistra in alto nelle due fotografie è composta di gas ionizzati. La coda piú debole piegata verso il basso è composta di polvere. La fotografia a sinistra è stata presa il 22 agosto, quella a destra cinque giorni dopo. Entrambe sono fatte col telescopio Schmidt di 120 centimetri posto sul monte Palomar.

La cometa Humason, una cometa visibile nel 1960, fu un caso unico in quanto il suo spettro non presentava niente polvere alla sua distanza al perielio di circa quattro unità astronomiche dal Sole. Il colore blu è dovuto all'emissione di molecole ionizzate come monossido di carbonio (CO^+), azoto molecolare (N_2^+), radicale ossidrile (OH^+), cianogeno (CN) e CH^+. Siccome il telescopio (lo Schmidt di 120 centimetri sul monte Palomar) seguiva il moto della cometa, le stelle appaiono come strisce.

piú centimetri si consumasse (sublimasse) mentre la cometa attraversava questo inferno solare. Un banco di ghiaia poco compatta avrebbe dovuto volatilizzarsi completamente quando le forze mareali dovute alla gravità solare lo strappavano via. L'intenso vento solare avrebbe dovuto poi allontanare i gas risultanti e completare la distruzione della cometa. La cometa Ikeya-Seki superò la prova praticamente indenne. In verità fa parte di un gruppo di otto di tali comete. Sembra che esse abbiano un antenato comune che si spezzò in comete separate, probabilmente a causa delle forze mareali dovute alla gravità solare.

Le comete radenti al Sole passano cosí vicino a esso da trovarsi entro la ben nota distanza limite stabilita dal francese E. Roche verso la metà del diciannovesimo secolo. Roche dimostrò che quando un piccolo corpo fluido supera un corpo grande entro due volte e mezzo il raggio del corpo maggiore, la forza gravitazionale sulla parte prossima del corpo minore è di molto superiore a quella sulla parte lontana. La forza è sufficiente a separare il corpo opponendosi alla propria semplice coesione gravitazionale. Solo un corpo che possegga piú della coesione gravitazionale può opporsi agli effetti mareali entro il limite di Roche. Persino un corpo con tale coesione può spaccarsi lungo screpolature o in altre zone di fragilità. Si può perciò concludere che i nuclei delle comete, almeno di quelle appartenenti alla famiglia delle comete radenti al Sole, posseggono una qualche forza interna che non sempre è però sufficiente a proteggerle dalla rottura mareale di un incontro particolarmente prossimo al Sole. Questi fatti sono sufficienti a eliminare il modello del banco di ghiaia del nucleo cometario a meno che non si desideri postulare che le varie comete siano completamente differenti nella loro struttura di base.

La seconda prova sperimentale

La seconda prova sperimentale fatale per il modello del banco di ghiaia è la persistenza di alcune comete periodiche. Circa 100 comete delle 600 finora osservate si muovono su orbite con periodo inferiore a 100 anni. Tutte quelle con periodo inferiore a 30 anni si muovono intorno al Sole nella stessa direzione dei pianeti e i piani delle loro orbite sono inclinati in media di 12 gradi rispetto al piano dell'orbita terrestre. Alcune di queste comete hanno completato molte rivoluzioni intorno al Sole dall'inizio dell'era storica. Le testimonianze diligenti dei cinesi mostrano che la cometa Halley è comparsa vicino al Sole 29 volte a intervalli tra 76 e 77 anni. Il primo passaggio ben osservato fu quello del 239 a.C. e il piú recente nel 1910. Vedremo di nuovo la cometa Halley nel 1986.

Dalle osservazioni fatte nello spazio dai satelliti « Orbiting Astronomical Observatory » (OAO) sappiamo che le comete brillanti perdono tonnellate di gas al secondo per periodi di mesi quando si approssimano al Sole. Secondo la teoria del banco di ghiaia ci si aspetterebbe un passaggio singolo per allontanare una frazione preponderante del gas che le particelle del banco di ghiaia hanno assorbito. Le comete dovrebbero quindi diminuire drasticamente di splendore ai passaggi successivi. La cometa Halley ha perso probabilmente un chilometro cubo di ghiaccio o forse anche di piú durante la parte nota della sua storia. Lo spazio interplanetario contiene cosí poca materia che il banco di ghiaia non potrebbe assorbire abbastanza gas per rinnovare il suo rifornimento durante la parte di vita spesa a grandi distanze dal Sole. Perciò una cometa deve avere un'enorme riserva di ghiaccio e materiali solidi per procurarsi il combustibile necessario alla sua spettacolare esibizione a ogni passaggio vicino al Sole. Per la cometa Halley un nucleo di ghiaccio di qualche decina di chilometri o piú di diametro è chiaramente sufficiente. La cometa deve perdere solo una frazione dell'un per cento della sua massa totale a ogni ritorno.

Arriviamo cosí alla prova finale. Una misteriosa idiosincrasia delle comete è quella che esse sfidano la legge gravitazionale di Newton. Nel 1819 il tedesco J.F. Encke studiò il moto di una cometa a breve periodo scoperta nel 1786. Con sua sorpresa trovò che la cometa, che ha il piú breve periodo conosciuto (3,3 anni), continuava a ritornare a ogni rivoluzione con due ore e mezzo di anticipo. Solo due ore e mezzo su un periodo di 3,3 anni può sembrare uno scarto insignificante, ma la precisione con cui si possono misurare le posizioni astronomiche la pone ben al di fuori dei possibili errori di misura e di calcolo.

Moto non gravitazionale

La cometa Encke continua a ritornare in anticipo rispetto al previsto a ciascuna rivoluzione. Brian G. Marsden e Zdenek Sekanina dello Smithsonian Astrophysical Observatory trovano però che l'effetto è lentamente diminuito per circa il 10 per cento della deviazione rivelata 150 anni fa. Anche la cometa Halley mostra tendenze individuali sin-

a

b

c

Le comete sfidano la legge di gravitazione accelerando e rallentando sulle loro orbite. La radiazione proveniente dal Sole fa sí che il ghiaccio sublimi ed espella molecole a velocità di parecchie decine di metri al secondo. Le molecole (*in colore*) sono espulse dalla parte soleggiata del nucleo e generano un effetto razzo. Questa forza spinge la cometa lontano dal Sole (*a*). Se il nucleo ruota nella stessa direzione del suo moto intorno al Sole (*b*), la reazione ritardata del getto spingerà avanti la cometa sulla sua orbita. Il nucleo sarà trasportato verso l'esterno del Sole, il suo periodo orbitale aumenterà e apparirà piú tardi del previsto. Se il nucleo ruota in direzione opposta al suo moto orbitale intorno al Sole (*c*), l'effetto razzo avrà una componente del moto opposta alla direzione di spostamento della cometa. La cometa sarà soggetta a una forza frenante e sarà trasportata verso l'interno in direzione del Sole. Il suo periodo orbitale si ridurrà ed essa comparirà prima del previsto.

golari nel suo moto. T. Kiang del Dunsink Observatory di Dublino trova che durante le ultime undici apparizioni la cometa continua a ritardare il suo arrivo di una media di 4,1 giorni. Marsden e Sekanina notano che delle 20 comete a breve periodo tutte, salvo due, presentano deviazioni dal moto newtoniano, metà in anticipo e metà in ritardo.

Il modello del banco di ghiaia non riesce assolutamente a spiegare questo moto non gravitazionale delle comete. Fino a quando le sonde spaziali non hanno dimostrato che lo spazio interplanetario è quasi perfettamente vuoto, si poteva ammettere che vi fosse un ostacolo dovuto a qualche mezzo resistente. Ora sappiamo però che nello spazio non vi è abbastanza materia per produrre l'effetto di ritardo richiesto. Inoltre, anche se vi fosse questo tipo di ostacolo, non avrebbe modo di far avanzare alcune comete sulle loro orbite.

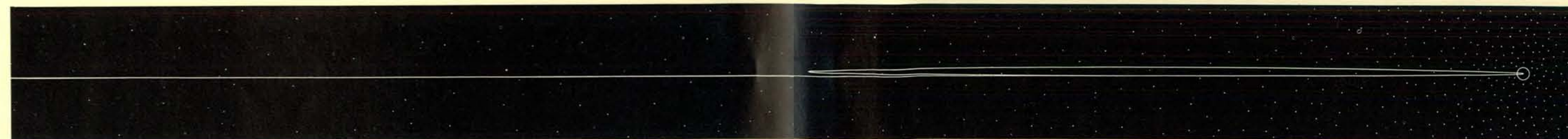
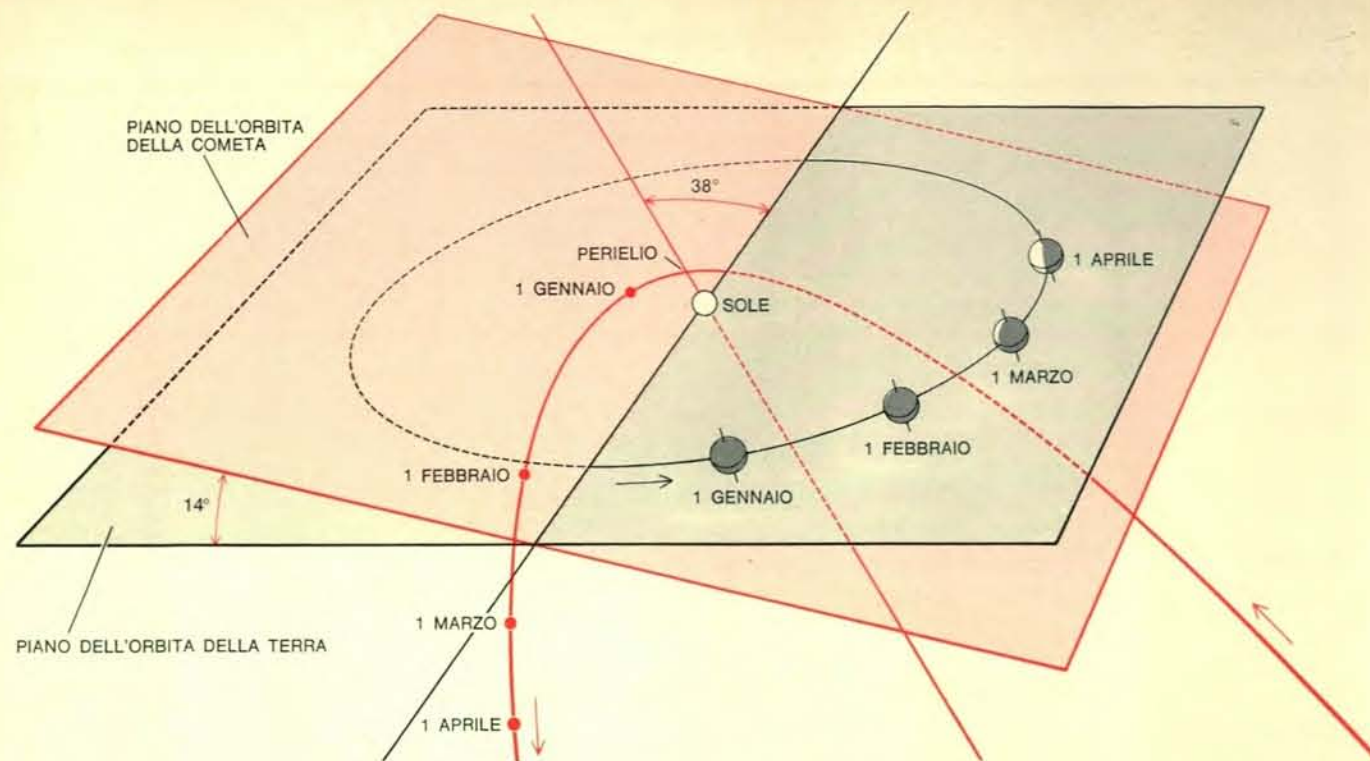
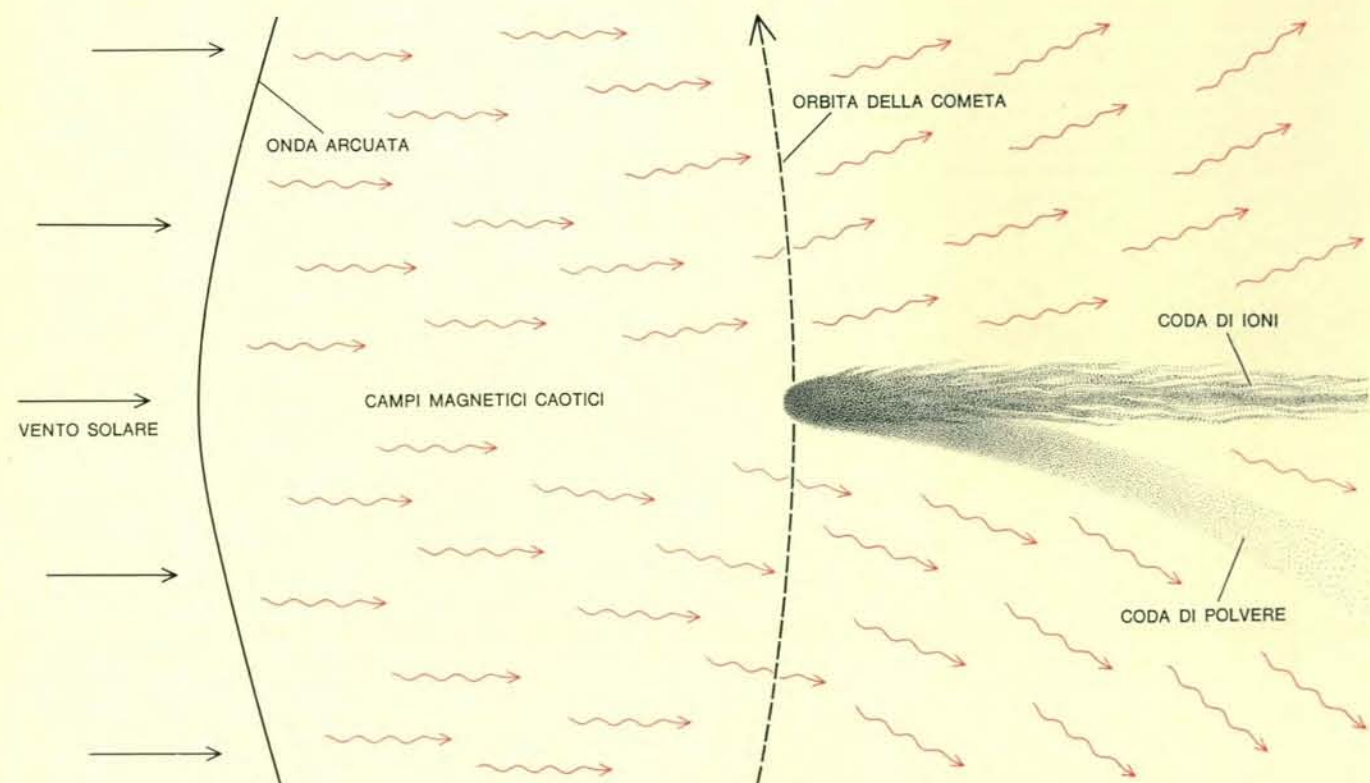
Il nucleo di ghiaccio fornisce una spiegazione semplice per il moto capriccioso delle comete. Nel fulgore della radiazione solare, il ghiaccio evapora e fa emergere molecole a velocità di parecchie decine di metri al secondo. Le molecole sono espulse dalla parte di nucleo rivolta verso il Sole e generano un effetto razzo, una reale forza che allontana la cometa dal Sole. Se però questa forza deve cambiare il periodo di una cometa, deve essere diretta lungo il percorso della cometa stessa sia in testa che in coda. Supponiamo ora che il nucleo della cometa stia ruotando (e dobbiamo ancora osservare un oggetto celeste che non ruoti in qualche misura). Il getto verrà spostato dalla rotazione, vi sarà cioè un ritardo nell'espulsione del gas da qualsiasi punto sulla superficie quando esso ruota oltre il punto direttamente opposto al Sole.

Il ritardo è simile a quello delle stagioni sulla Terra: la temperatura estiva nell'emisfero boreale è massima alla fine di luglio e non al momento del massimo riscaldamento solare il 21 giugno. Se una cometa sta ruotando nella stessa direzione del suo moto intorno al Sole, l'azione ritardata di razzo avrà una componente in avanti. La cometa sarà trascinata lontano dal Sole sulla sua orbita, il suo periodo orbitale aumenterà e come risultato apparirà piú tardi del previsto.

D'altra parte, se una cometa ruota in direzione opposta alla direzione del suo moto intorno al Sole, l'effetto razzo avrà una componente all'indietro. La cometa sarà sottoposta a una forza di resistenza e verrà trascinata lentamente verso il Sole sulla sua orbita. I calcoli che ho fatto nel 1950 dimostrano che per un nucleo cometario che ha un diametro dell'ordine di pochi chilometri e che utilizza con una certa efficienza l'energia solare per far evaporare i suoi ghiacci, la forza calcolata è sufficiente per produrre le variazioni osservate nei tempi di comparsa delle comete periodiche. Se gli assi di rotazione delle comete sono distribuiti a caso ci aspetteremmo che metà di esse siano accelerate e metà ritardate, ciò che si verifica nella realtà.

La componente acqua

Ho trovato successivamente che, se il modello del nucleo di ghiaccio deve rendere conto delle variazioni di periodo delle comete, dovrebbe espellere gas a un tasso considerevolmente piú alto di quello che si può dedurre dall'intensità delle righe dello spettro di una cometa alle lunghezze d'onda della luce visibile. In caso contrario, l'effetto razzo non sarebbe sufficientemente forte per esercitare un'adeguata forza su nuclei maggiori. Biermann dimostrò allora che se la normale acqua ghiacciata è la principale componente dei nuclei di ghiaccio, le comete brillanti devono essere circondate da enormi nubi di idrogeno e radicali ossidrile che provengono da molecole di acqua dissociata. Un atomo di idrogeno, isolato nella immensità dello spazio, emette quasi tutta la sua radiazione fluorescente non nella regione visibile dello spettro, ma nell'ultravioletto assai lontano in una riga spettrale chiamata Lyman alfa.

Nel 1970 Arthur D. Code e Charles F. Lillie puntarono il primo Orbiting Astronomical Observatory verso la cometa Bennett. Scoprirono una vasta nube di idrogeno che emetteva alle lunghezze d'onda della riga Lyman alfa. Trovarono anche che la radiazione dell'ossidrile nel lontano ultravioletto, che è appena visibile sulla Terra, è molto piú intensa di quanto si fosse sospettato. Il gas che emetteva radiazione a queste lunghezze d'onda era quello che mancava nei miei calcoli. Sembra inoltre che sia approssimativamente uguale il numero di atomi di idrogeno e di radicale ossidrile, confermando il sospetto che l'acqua gelata sia la principale componente di un nucleo cometario. Cosí le osservazioni confermarono l'ipotesi di Biermann e le

Le comete, secondo la teoria proposta da J. H. Oort, provengono da un'enorme nube di nuclei staccati di ghiaccio. La nube di Oort gravitazionalmente fa parte del sistema solare, ma si estende per un raggio di circa 50 000 unità astronomiche (l'unità astronomica è la distanza media tra la Terra e il Sole). In questa sezione della nube di Oort, il cerchio a destra rappresenta il sistema solare; i puntini, i nuclei cometari. Una stella che attraversa le regioni esterne della nube (circa tre metri a sinistra fuori della pagina nella scala del disegno) disturberà il moto di alcuni nuclei abbastanza perché entrino nel sistema solare. L'ellisse rappresenta in scala corretta l'orbita della cometa Kohoutek: larga 44 unità astronomiche e lunga 3600.

L'orbita della cometa Kohoutek (*in colore scuro*) è disegnata rispetto all'orbita terrestre (*in nero*). L'orbita della Kohoutek è un'ellisse estremamente allungata, mentre l'orbita terrestre è quasi circolare. Il piano dell'orbita della cometa (*in colore chiaro*) è inclinato di un angolo di soli 14 gradi sul piano dell'orbita terrestre (*in grigio*). Al perielio la cometa era a circa 21 milioni di chilometri dal Sole. Le date lungo le due orbite indicano le rispettive posizioni della cometa e della Terra.

Le code di ioni e di polveri di una cometa si creano per due diversi processi. La coda di ioni è formata a sua volta per un doppio processo. Primo, gli elettroni ad alta energia del vento solare (gas espulso dal Sole ad alta velocità) ionizzano le molecole della chioma cometaria strappando loro elettroni e lasciandole cariche positivamente. Secondo, il vento solare dà origine a un'onda arcuata intorno alla chioma. I campi magnetici caotici entro il vento solare agiscono come rastrelli magnetici che allontanano selettivamente le molecole ionizzate dalla chioma ad alta velocità. La coda di polvere è dovuta alla pressione della luce solare che spinge fuori dalla chioma le particelle di polvere. La coda è curva perché le particelle di polvere seguono le leggi del moto dei corpi orbitanti e ritardano rispetto alla chioma dato che si muovono a velocità relativamente basse.

conclusioni fondamentali tratte dal modello del nucleo di ghiaccio. Le osservazioni furono confermate da J. Blamont e J. Bertaux del Centre National des Etudes Spatiales di Parigi; da allora sono state scoperte grandi quantità di idrogeno in tre altre comete compresa la cometa Kohoutek.

L'origine delle comete

Ora che conosciamo la natura di un nucleo cometario, quale ne sarà l'esatta struttura chimica e fisica? Questa struttura può svelare l'origine, l'evoluzione, il tempo di vita e la fine delle comete? È facile fare comete almeno stando seduti a tavolino. Abbiamo solo bisogno di collassare e raffreddare una nube di gas e polvere interstellare. Nel piano della nostra galassia esistono molte nubi collassate in cui stanno ora evolvendo nuove stelle e probabilmente nuovi sistemi solari. Una di queste incubatrici stellari è la regione della grande nebulosa di Orione.

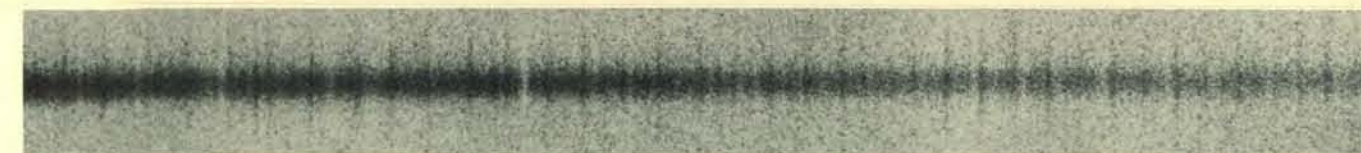
Quando in una nube molto collassata la temperatura scende in maniera sufficiente, per esempio a poche decine di gradi sopra lo zero assoluto, tutte le sostanze terrestri come acqua, ammoniaca e metano congelano in particelle solide. Le molecole che si osservano nelle comete sono semplici combinazioni di carbonio, azoto, ossigeno e idrogeno. Se le comete hanno origine in una tipica nube interstellare della nostra galassia dove le abbondanze atomiche sono assai simili a quelle solari, questi quattro elementi sono di gran lunga quelli preponderanti, capaci di formare composti. Sebbene un quarto della massa solare sia costituito di elio, l'elio e gli altri gas nobili non si combinano chimicamente con altri elementi e non congelano alle temperature e alle densità previste.

Le comete che osserviamo si formano realmente mediante questi processi? Le comete del sistema solare formano una enorme nube che gravitazionalmente fa parte del sistema solare, ma che si estende parecchie migliaia di volte più lontano del pianeta più esterno (Plutone). Come ha dimostrato J.H. Oort dell'Università di Leida nel 1950, una stella che passa vicino al limite esterno di questa nube cometaria disturberà il moto di alcune comete in maniera sufficiente perché esse entrino nella parte interna del sistema solare dove noi possiamo osservarle. Altrimenti restano a bassissime temperature, forse fin dall'epoca in cui si sono formati il Sole e i pianeti: 4,6 miliardi di anni fa. Sembra che la cometa Kohoutek sia una delle comete « nuove » che entrano per la prima volta nella regione planetaria; potrebbe perciò essere un campione della materia primordiale del sistema solare.

Nube interstellare contro nube solare

Dove si sono formate le comete nella grande nube di Oort? A.G.W. Cameron dell'Harvard College Observatory ha recentemente suggerito che esse si siano formate da nubi interstellari frammentarie quando una nube di proporzioni maggiori si appiattì in un disco per formare il sistema solare. Le comete in quel momento si sarebbero trovate su enormi orbite, simili a quelle delle nuove comete, come la cometa Kohoutek, prima di venire disturbate da una stella di passaggio. Non è ancora sicuro se questi ghiacci e solidi necessari alla formazione di comete piuttosto grandi possono aggregarsi in una nube interstellare frammentaria. D'altra parte siamo sicuri che le comete si devono essere formate in enormi quantità ai limiti esterni del sistema solare.

I pianeti Urano e Nettuno hanno una densità media circa doppia di quella dell'acqua, proprio quella che ci si attenderebbe se fossero letteralmente aggregati di centinaia di miliardi di comete. I pianeti terrestri (Terra, Marte, Venere e Mercurio) devono essersi formati da un gran numero di planetesimi, o piccoli aggregati di materia simile a quella terrestre. La differenza tra Urano e Nettuno, da una parte, e i pianeti terrestri, dall'altra, sembra che dipenda semplicemente dalla temperatura nelle due regioni di spazio al tempo della loro formazione. Ci si aspetterebbe che in una grande nube di materia interstellare collassata, che ruota come un disco, la temperatura debba restare maggiore vicino al protosole, o Sole in formazione, che non ai bordi del disco. Evidentemente l'acqua non poteva gelare entro l'orbita di Giove, ma poteva gelare fuori dell'orbita di Saturno. (Giove e Saturno stessi, come il Sole, sono composti essenzialmente di idrogeno e elio compressi dalla grande massa dei pianeti a densità prossime a quelle dell'acqua.) Perciò possiamo essere abbastanza sicuri che le comete si sono formate oltre l'orbita di Saturno. I grani di ghiaccio e quelli di materia simile a quella terrestre, for-

Lo spettro di una cometa che ha molto gas rispetto alla quantità di polvere mostra molte righe spettrali distinte che compongono una banda molecolare. Lo spettro, rappresentato in questa figura, è quello della cometa Ikeya (visibile nel 1963); esso abbraccia le lunghezze d'onda blu da 3900 angstrom (*bordo a sinistra*) a 4100 angstrom (*bordo a destra*). Dato che questo spettro è una stampa in negativo, le righe scure sono in realtà brillanti righe di emissione. Quasi tutte le righe sono dovute alla molecola del carbonio C_3. La striscia orizzontale priva di particolari nel centro è il debole spettro della luce solare riflessa.

Lo spettro di una cometa che ha molta polvere rispetto ai gas mostra poche e deboli bande distinte e righe spettrali. La linea nera orizzontale dello spettro continuo della luce riflessa è però molto forte. Lo spettro, che è quello della cometa Mrkos, è anche solcato da righe di assorbimento che appaiono chiare sulla stampa in negativo. Queste indicano le lunghezze d'onda a cui la polvere assorbe luce dallo spettro solare. Lo spettro copre approssimativamente la regione tra 4880 (*bordo a sinistra*) e 5100 angstrom (*bordo a destra*). Gli spettri sono di J.L. Greenstein dell'Istituto di tecnologia della California.

La chioma della cometa Bennett, una cometa brillante visibile durante la primavera del 1970, mostra una struttura a spirale ricca di dettagli in questa fotografia fatta da R. B. Minton e S. M. Larson il 28 marzo 1970. Essi fecero 10 fotografie con tempi di esposizione tra 4 e 8 secondi con il telescopio riflettore di 152 centimetri dell'Osservatorio Catalina. Questa figura è una composizione delle 10 immagini. Minton e Larson interpretano la struttura a spirale come prova che il nucleo di ghiaccio ruota intorno al suo asse con un periodo di circa un giorno e mezzo, mentre ruota intorno al Sole.

matisi nel disco in via di raffreddamento della nube solare, si accumularono in comete e finalmente si aggregarono in Urano e Nettuno. Dopo che Urano e Nettuno divennero massicci, le comete residue furono disturbate dall'attrazione gravitazionale sia dei nuovi pianeti sia di Giove e Saturno e furono costrette a ruotare sulle orbite enormi della nube di Oort.

Il processo dovrebbe essere stato enormemente dispendioso. Molte comete dovrebbero essere state catturate da Giove e Saturno, molte dovrebbero essere decadute in orbite piú piccole del sistema solare piú interno dove sono ora, molte dovrebbero essere andate perdute nella profondità della Galassia. Nessuno ha ancora dimostrato però che le comete formatesi entro le orbite degli attuali pianeti esterni potrebbero veramente muoversi al di fuori, nella grande nube di Oort, né, come abbiamo visto, è stato dimostrato che nubi interstellari frammentarie possano condensarsi e aggregarsi in comete.

La soluzione del dilemma

Per determinare se le comete si formarono nella nube che diede origine al sistema solare, nelle nubi interstellari frammentarie o in qualche altro tipo di dintorni, sono necessari indizi forniti sia da nuove teorie sulle comete sia da nuove osservazioni. Supponiamo che le comete si siano formate in nubi interstellari frammentarie. Allora ciascun grano di polvere interstellare costituirebbe un nucleo per la condensazione dei vari ghiacci. I ghiacci possono consistere in larga misura di molecole «strane» ora scoperte dai radioastronomi nelle nubi di gas e polvere interstellari. Queste molecole comprendono formaldeide (CH_2NH), alcool metilico (CH_3OH), cianuro di metile (CH_3CN), cianuro di idrogeno (HCN) e piú di 20 altri composti per lo piú di elementi abbondanti nelle comete: carbonio, azoto, ossigeno e idrogeno.

Alcune di queste molecole sono le molecole che hanno generato i radicali semplici che osserviamo nelle comete? Se lo sono, allora le comete si sono probabilmente formate in nubi interstellari frammentarie. Se, d'altra parte, le comete hanno avuto origine entro le orbite dei pianeti esterni, ci si aspetterebbe che le molecole originali siano quelle piú stabili come ammoniaca e metano. Come hanno suggerito Pol Swings e Armand Delsemme dell'Istitut d'Astrophysique di Liegi, anche se la temperatura della nube solare non può mai essere stata tanto bassa da far gelare il metano (circa 20 gradi sopra lo zero assoluto), queste molecole furono probabilmente accalappiate dalla neve di acqua.

I due possibili luoghi di nascita delle comete possono anche dare origine a marcate differenze nella struttura e nel comportamento delle comete. Ci si potrebbe attendere che le comete della nube solare fatte di grani di polvere circondati da ghiacci si disintegrano dando null'altro che i prodotti di sublimazione delle loro sostanze. D'altra parte le comete che si sono formate nella nube solare esterna possono avere ben sviluppato un nucleo simile a quello terrestre se la temperatura del grande disco della nube al principio aumentò, durante il suo collasso, e poi scese lentamente. Una cometa vecchia di questo tipo può perdere il suo involucro esterno di ghiaccio, diventare lentamente inattiva e apparire infine con un aspetto stellare e non con un aspetto confuso nei nostri telescopi, cioè con un aspetto assai simile a quello di un pianetino. Infatti molti astronomi ritengono che i pianetini piú piccoli che osserviamo tagliare l'orbita terrestre siano in realtà vecchi nuclei cometari.

Il modo migliore per rispondere a queste domande, fondamentali per la comprensione dell'origine del sistema solare, è quello di inviare sonde spaziali verso il nucleo di una cometa e verso la superficie di un pianetino. Queste sonde spaziali, che sarebbero in realtà osservatori senza uomini a bordo, sono oggi perfettamente realizzabili sia da un punto di vista tecnico sia da un punto di vista scientifico e sarebbero meno costose di molte sonde planetarie che hanno già dato un esito tanto positivo.

La NASA sta attualmente studiando la possibilità di tali missioni spaziali verso comete e pianeti. La prima missione che si sta oggi prendendo in esame sarà per la cometa Encke nel 1980. Intanto l'interesse per le comete, rinnovatosi in occasione del passaggio della cometa Kohoutek, dovrebbe tradursi in nuove informazioni di fondamentale importanza per la comprensione della vera natura dei corpi piú primitivi del sistema solare.

Nutrizione e cervello

Una dieta ricca di carboidrati fa aumentare nel cervello la produzione di serotonina. Il meccanismo potrebbe far parte di un circolo chiuso nel quale la dieta influisce sul consumo di cibo

di John D. Fernstrom e Richard J. Wurtman

Il cibo influenza la funzione cerebrale? Per porre la domanda in modo piú specifico, i cambiamenti nella chimica del sangue conseguenti all'ingestione di sostanze nutritizie producono cambiamenti corrispondenti nei tessuti del cervello? Se cosí è, questi cambiamenti indotti dalla dieta possono influenzare l'attività funzionale del cervello? Infine, potrebbe il cervello «affamato» nutrirsi da solo «ordinando» al corpo di nutrirsi?

Nel passato la risposta generale a questa domanda è stata no. Il cervello estrae l'ossigeno, il glucosio e altre sostanze nutritizie, come gli amminoacidi dal sangue a una velocità che sembra non avere correlazione con la concentrazione di queste sostanze nel sangue. Per questa ragione non si è pensato che le fluttuazioni passeggere nella composizione del sangue in seguito all'ingestione di cibo o durante un periodo di digiuno influissero sul cervello. Vi è, tuttavia, un'eccezione a questa regola generale e noi e i nostri colleghi del Dipartimento della nutrizione e delle scienze alimentari al Massachusetts Institute of Technology l'abbiamo studiata per molti anni.

L'eccezione è la seguente. Dopo un pasto ricco in carboidrati si ha un aumento nella velocità alla quale il cervello sintetizza la serotonina, un mediatore chimico del sistema nervoso. I mediatori del sistema nervoso sono le sostanze, presenti nei neuroni dei mammiferi, che, quando vengono liberate, trasmettono segnali attraverso le sinapsi ad altri neuroni all'interno del cervello o a cellule muscolari o ancora a cellule secretorie al di fuori del cervello. La serotonina è uno dei sei composti dei quali si è stabilito con certezza sufficiente la funzione di mediatori nel cervello dei mammiferi. Gli altri sono l'acetilcolina, l'acido gamma-amminobutirrico e le tre catecolammine: l'adrenalina, la noradrenalina e la dopammina.

Nei ratti di laboratorio la risposta del cervello all'ingestione di carboidrati comincia entro un'ora dopo il pasto. È preceduta da un cambiamento nella concentrazione normale di amminoacidi nel sangue. La concentrazione della maggior parte degli amminoacidi diminuisce, ma la concentrazione di un amminoacido, il triptofano, aumenta. Come conseguenza si ha un aumento proporzionale nell'assorbimento di triptofano da parte del cervello. Dato che il triptofano è un precursore nella sintesi della serotonina, dopo l'ingestione di carboidrati i neuroni del cervello che convertono il triptofano in serotonina sono in grado di farlo a una velocità maggiore, e il livello di serotonina nel cervello aumenta. Poiché probabilmente viene liberata una quantità superiore di mediatore ogni volta che i neuroni contenenti serotonina vengono eccitati può darsi che la funzione del cervello ne sia modificata.

Il fatto che il cervello debba essere sensibile a cambiamenti del sangue indotti dalla dieta sembra a prima vista non solo sorprendente ma addirittura sconcertante. Quale può essere il vantaggio fisiologico per l'individuo quando un organo dell'importanza del cervello può essere influenzato dai capricci della dieta? Se, in realtà, non vi è nessun vantaggio, come è possibile che un meccanismo aperto di questo tipo il quale modifica la funzione cerebrale abbia potuto sopravvivere alle prove della selezione naturale? Qualunque possa essere la risposta, non si può sfuggire alla conclusione che l'assunzione di sostanze nutritizie altera la quantità di serotonina presente nei neuroni serotoninergici del cervello.

Come le tre catecolammine, la serotonina è una monoammina; per essere precisi, si tratta della 5-idrossitriptammina. È stata scoperta per la prima volta nel 1948 da tre studiosi della clinica di Cleveland, Maurice M. Rapport, Arda Green e Irvine H. Page. Essi hanno inizialmente isolato la sostanza dalle piastrine del sangue di mammiferi. Poco dopo si apprese che quantità relativamente grandi di serotonina sono presenti nel cervello dei mammiferi, e in dieci anni i neuroanatomisti svedesi Annica Dahlström e Kjell Fuxe, lavorando con il microscopio a fluorescenza, hanno localizzato i neuroni nei quali è contenuta la serotonina. Essi hanno trovato che circa tutta la serotonina presente nel cervello è confinata in un gruppo distinto di neuroni conosciuti come i nuclei del rafe. I corpi cellulari di questi neuroni sono situati nel tronco encefalico; le loro fibre ascendono verso il resto del cervello e discendono attraverso il midollo spinale per formare sinapsi largamente distribuite.

Il triptofano è uno degli amminoacidi «essenziali», ossia non può essere sintetizzato dalle cellule dei mammiferi. Qualunque sia la riserva di triptofano di un animale, essa può essere ottenuta solamente ingerendo una proteina che contiene la sostanza. Non è questo il caso di molti altri amminoacidi, per esempio la tirosina, il precursore della noradrenalina e della dopammina. La tirosina può essere sintetizzata nel fegato dei mammiferi, con la fenilalanina, un altro amminoacido, come precursore. Il triptofano non solo dev'essere ottenuto per via esogena; esso è anche l'amminoacido meno abbondante nelle proteine della dieta; ne consegue quindi che l'apporto di triptofano al corpo dei mammiferi è di solito piuttosto piccolo.

La trasformazione di triptofano e tirosina in mediatori ha inizio nello stesso modo: con l'aggiunta di un gruppo idrossile (OH) attraverso l'azione catalitica di un enzima. I due enzimi coinvolti tuttavia differiscono in maniera significativa. La triptofano-idrossilasi è un enzima a bassa affinità. Ciò significa che una molecola di triptofano e una molecola dell'enzima non hanno grande propensione a combinarsi e, quindi, solamente quando la concentrazione di triptofano è molto superiore al normale, l'enzima può funzionare alla massima velocità. Al contrario, la tirosina-idrossilasi è un enzima ad alta affinità; esso opera con un'efficienza massima anche quando la concentrazione di tirosina non è superiore al normale.

Per la prima volta, nel 1968, ci siamo interessati alla relazione fra nutrizione e sintesi di serotonina, quando uno studente del nostro laboratorio, William Shoemaker, trovò che i cervelli di giovani ratti, sottoposti a una dieta carente di proteine, contenevano livelli anormalmente bassi di dopammina e di noradrenalina. Considerato che il triptofano è presente solo in piccole quantità anche in una dieta normale e, al contrario del precursore delle due catecolammine, è possibile ottenerlo solamente dalle proteine, ci sembrò che la dieta povera in triptofano avrebbe dovuto deprimere ancora di piú il livello della serotonina cerebrale.

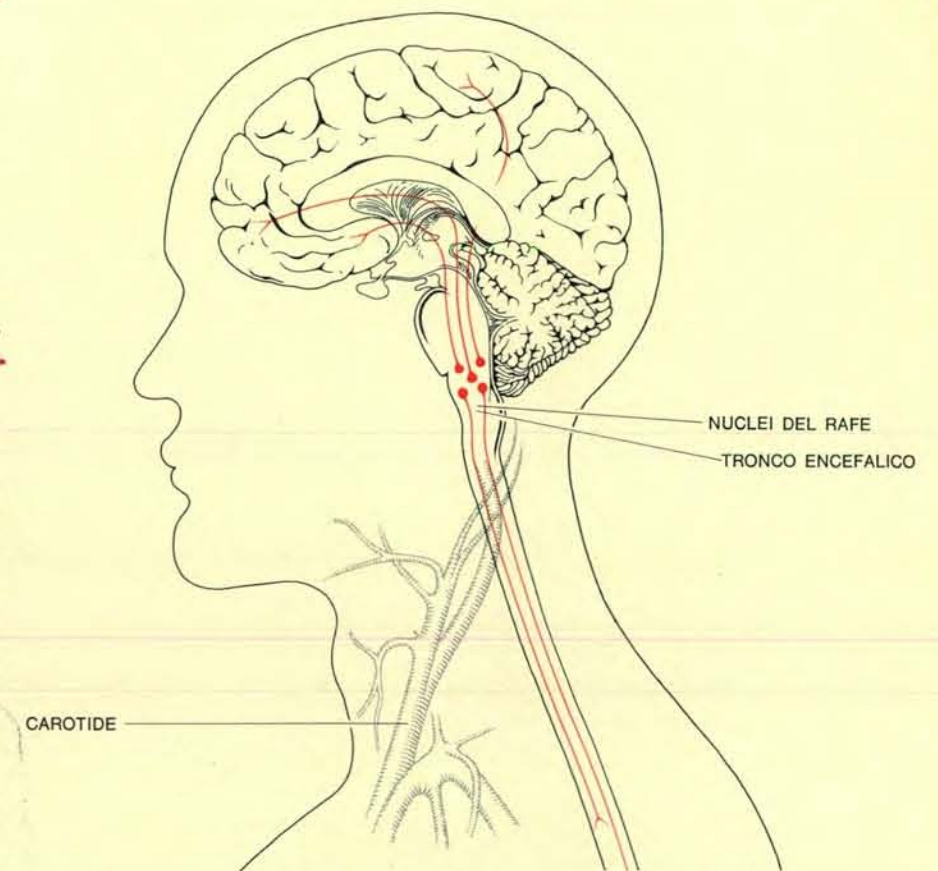
Abbiamo nutrito dei ratti per molte settimane con una dieta che aveva una sola fonte di proteine: il granoturco, un cibo conosciuto per essere particolarmente povero in triptofano. Quando abbiamo misurato i livelli di triptofano e di serotonina nei cervelli dei ratti, questi si rivelarono notevolmente al di sotto del normale. Avendo cosí stabilito che a lunga scadenza i livelli di serotonina cerebrale dipendono dai livelli di triptofano cerebrale, ci siamo chiesti se si poteva dimostrare la stessa dipendenza quando il cambiamento fosse rapido, a breve scadenza. Sembrava possibile che questi cambiamenti intensi potessero accadere spontaneamente nel corso di una giornata; avevamo osservato precedentemente che il livello di triptofano ematico nell'uomo aumenta e diminuisce secondo un ritmo quotidiano caratteristico.

Lavorando coi ratti, abbiamo trovato che gli animali sottoposti a una dieta normale e controllati a intervalli di quattro ore giorno e notte rivelavano anch'essi dei cambiamenti ritmici nel livello del triptofano ematico. Inoltre, esisteva anche un ritmo del triptofano cerebrale parallelo al ritmo del triptofano ematico.

Mentre stavamo compiendo queste ricerche, abbiamo saputo di un nuovo metodo di dosaggio chimico per la serotonina di gran lunga piú sensibile dei metodi fino a quel momento a disposizione. Usando questo nuovo metodo, abbiamo paragonato le concentrazioni di serotonina cerebrale e di triptofano cerebrale in gruppi di ratti sottoposti alla stessa dieta dell'esperimento precedente. Dato che il livello di serotonina nel cervello dovrebbe dipendere tanto dalla sua velocità di utilizzazione quanto dalla sua velocità di sintesi, non ci aspettavamo che i ritmi fossero perfettamente in fase. In realtà, non lo erano, ma erano tuttavia abbastanza vicini. Un leggero aumento pomeridiano nella concentrazione di triptofano cerebrale non era accompagnato da un aumento corrispondente nella concentrazione di serotonina cerebrale; in realtà la serotonina cerebrale diminuiva. In seguito, tuttavia, i due ritmi erano generalmente in fase.

Abbiamo completato questi studi sul ritmo nel 1970 ed eravamo convinti che nei ratti le variazioni naturali nelle concentrazioni di triptofano cerebrale influenzassero la sintesi di serotonina da parte del cervello. Dopo poco tempo abbiamo iniziato una ricerca bibliografica sui fattori che influiscono sulla sintesi di altri mediatori. Avendo compiuto ricerche su cambiamenti quotidiani, siamo stati piuttosto sorpresi di trovare che la maggior parte degli altri esperimenti si erano occupati solo di cambiamenti a lungo termine. Il nostro punto di vista era che, se la disponibilità dei precursori era un fattore fisiologico significativo che influenzava la sintesi del mediatore, allora i cambiamenti rapidi, a breve scadenza, nel livello delle concentrazioni del precursore sarebbero stati piú importanti. Abbiamo deciso di vedere se un aumento artificiale rapido del livello del triptofano cerebrale nei ratti, prodotto da un'iniezione dell'amminoacido, avrebbe aumentato la concentrazione cerebrale di serotonina.

Le nostre prime iniezioni erano sufficientemente grandi da fornire al ratto una quantità corrispondente circa alla metà della dose di triptofano contenuta nella razione di cibo quotidiano. Entro un'ora la concentrazione di triptofano nel cervello del ratto raggiungeva un livello nove volte superiore a quello normale e la concentrazione di serotonina quasi raddoppiava. Anche quando riducevamo la quantità di triptofano iniettata a un decimo della dose, i livelli di triptofano cerebrale e di serotonina aumentavano ancora.

Avendo dimostrato che anche piccoli aumenti nel triptofano cerebrale causano un aumento della concentrazione di serotonina cerebrale, abbiamo deciso di vedere se una diminuzione nel livello del triptofano cerebrale avrebbe avu-

I neuroni cerebrali, nei quali è contenuta la maggior parte di serotonina, un mediatore chimico del sistema nervoso, costituiscono un gruppo conosciuto come nuclei del rafe. I corpi cellulari dei neuroni si trovano nel tronco encefalico; le fibre nervose ascendono verso il resto del cervello e discendono attraverso il midollo spinale.

to l'effetto di diminuire la velocità di sintesi di serotonina da parte del cervello. È noto che le iniezioni di insulina abbassano la concentrazione di amminoacidi nel sangue dei mammiferi e dei rettili, e così ci siamo accinti a provare la possibilità di abbassare il livello di serotonina cerebrale nei ratti iniettando loro dell'insulina. Lungi dall'abbassare la concentrazione di triptofano nel sangue, l'iniezione l'aumentava. Due ore dopo l'iniezione, la concentrazione di triptofano ematico raggiungeva un massimo e i livelli di triptofano e di serotonina cerebrali erano aumentati in modo simile. Il risultato ottenuto non era solo paradossale; era anche la prima prova sperimentale che un ormone (in questo caso l'insulina) può influenzare la sintesi di un mediatore nel cervello.

Ma era corretta la nostra analisi dell'esperimento? Forse l'aumento della serotonina cerebrale non era in relazione con l'aumento del triptofano cerebrale. Avrebbe potuto dipendere da cambiamenti di natura riflessa nell'attività neuronale causati dalla diminuzione precipitosa nel livello di zucchero ematico indotta dall'iniezione di insulina. Per provare questa possibilità abbiamo deciso di permettere ai ratti stessi di introdurre l'insulina nel loro sangue. Avremmo permesso ai ratti di seguire una dieta ricca in carboidrati, e perciò di stimolare il pancreas a secernere naturalmente l'insulina.

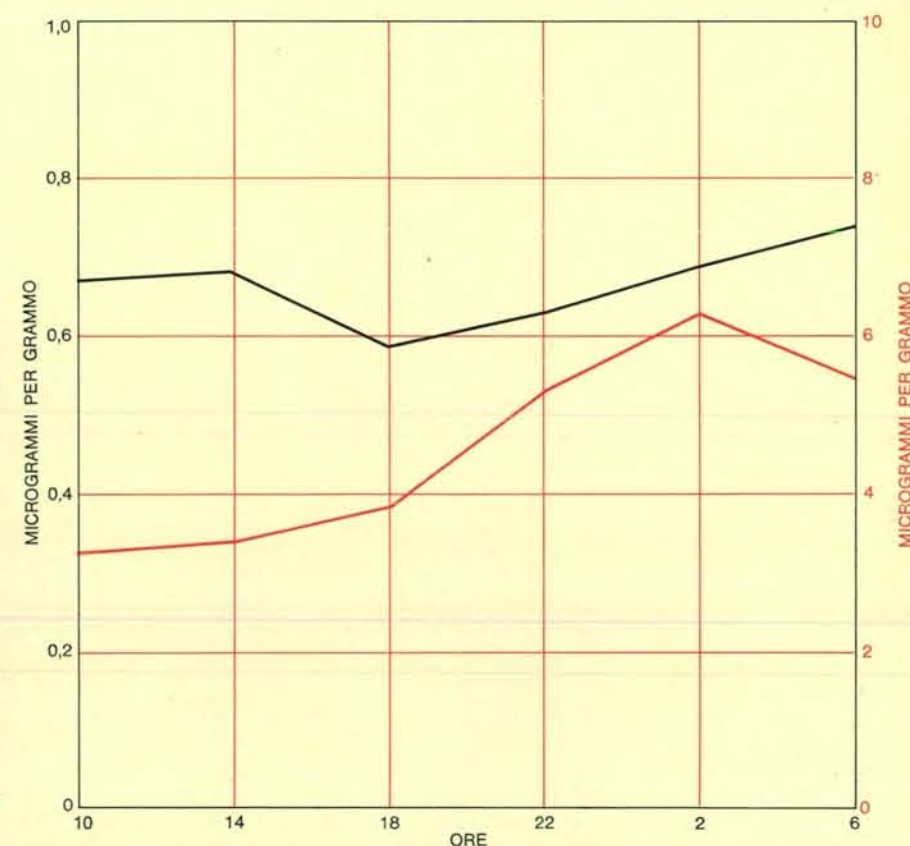
Abbiamo fatto digiunare gruppi di ratti per una notte e abbiamo poi dato loro una razione contenente carboidrati e grassi ma senza proteine. La secrezione naturale di insulina che ne seguì produsse gli stessi effetti delle iniezioni di insulina: nel giro di un'ora i livelli di triptofano nel sangue salirono notevolmente e raggiunsero il massimo in due ore. Durante la prima ora il livello del triptofano cerebrale aumentò più del 20 per cento e dopo due ore raggiunse un massimo del 65 per cento sopra il normale. La concentrazione di serotonina cerebrale aumentò nel corso della prima ora ed era circa del 20 per cento al di sopra del normale dopo due ore (*si veda la figura a pagina 73*).

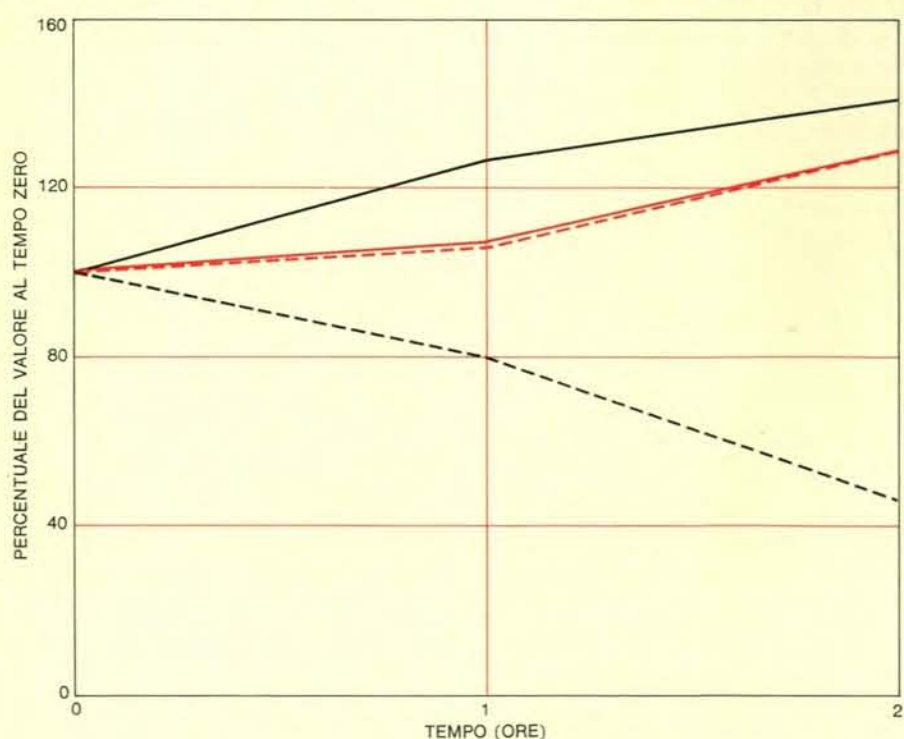
Rivedendo i nostri risultati abbiamo raggiunto due conclusioni e azzardato una previsione. Primo, la liberazione naturale di un ormone (l'insulina) in risposta al consumo di carboidrati aumenta il livello di un mediatore del sistema nervoso (serotonina) nel cervello del ratto; evidentemente è l'aumento nelle concentrazioni di triptofano cerebrale che accelera la sintesi di serotonina da parte del cervello. Inoltre, non solo questo stimolo «naturale» dell'insulina ma anche l'apporto artificiale di insulina o di triptofano mediante un'iniezione causa i medesimi effetti: un aumento del livello di triptofano nel sangue, con un aumento parallelo nei livelli di triptofano e di serotonina nel cervello. Sulla base di queste conclusioni abbiamo previsto che qualunque aumento nel livello del triptofano ematico sarebbe seguito da un aumento nella concentrazione della serotonina cerebrale. Per esempio, se sottoponiamo i ratti a una dieta che comprende proteine (la fonte naturale del triptofano), carboidrati e grassi, ci aspetteremmo che i livelli del triptofano cerebrale e della serotonina cerebrale aumentino in proporzione all'aumento del triptofano ematico.

Abbiamo verificato la nostra previsione sottoponendo due gruppi di ratti a diete speciali. A un gruppo abbiamo somministrato una razione di carboidrati e di grassi integrata per il 18 per cento con una proteina naturale del latte, la caseina. All'altro gruppo abbiamo somministrato la caseina con un miscuglio completo di amminoacidi.

Sono stati necessari soltanto un paio di pasti per dimostrare che la nostra previsione era sbagliata. Con tutte e due le diete integrate i livelli di trip-

La serotonina, la dopammina, la noradrenalina e l'adrenalina sono quattro mediatori che si trovano nel cervello dei mammiferi (*a, a sinistra*). Il precursore della serotonina è il triptofano, un amminoacido (*c, a destra*); il primo passo della sintesi è catalizzato (*freccia in colore*) da un enzima, la triptofano-idrossilasi. Il triptofano, un amminoacido «essenziale», viene fornito solo dalle proteine della dieta. Gli altri mediatori (*b, al centro*) sono formati in stadi successivi dallo stesso precursore, la tirosina, un amminoacido; la reazione iniziale (*freccia in colore*) è catalizzata da un enzima, la tirosina-idrossilasi. La tirosina, non essendo un amminoacido essenziale, può essere sintetizzata nel fegato dei mammiferi.

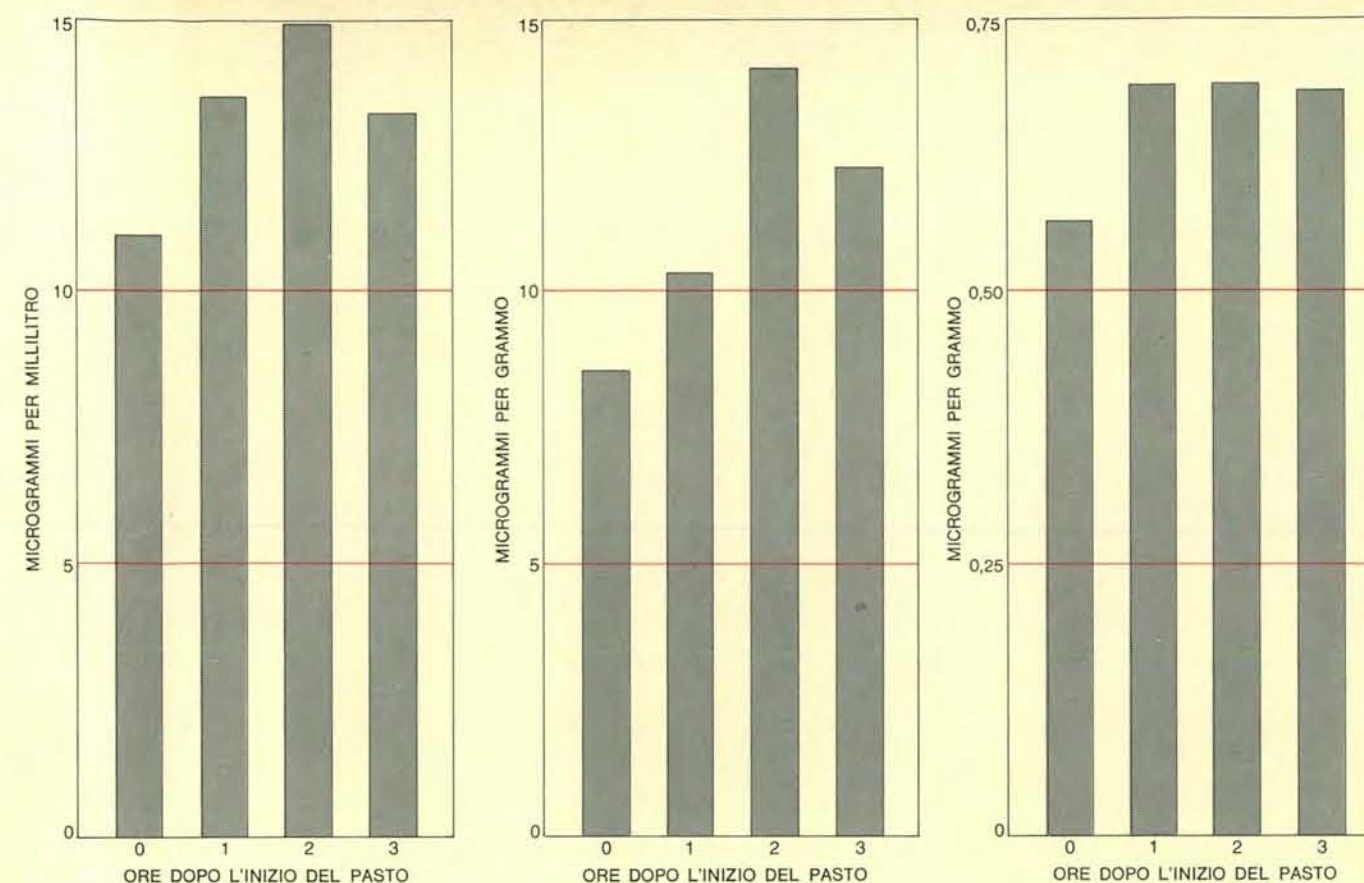
Fluttuazione quotidiana nel livello di triptofano nel sangue di ratti di laboratorio (*curva in nero*). Il ritmo era in relazione all'ora dell'assunzione del cibo. Anche il livello del triptofano cerebrale fluttuava (*curva in colore*); i ritmi erano in fase.

Nel cervello dei ratti fluttua anche la concentrazione di serotonina (*curva in nero*). Il ritmo che riflette due processi interagenti, l'accumulazione del mediatore sintetizzato dal triptofano e la sua liberazione nella scarica neuronale, non è proprio in fase con quello del triptofano cerebrale (*curva in colore*), ma risulta evidente una correlazione.

Iniezioni di diverse quantità di triptofano provocarono un rapido aumento nel livello di triptofano cerebrale (*colonne in colore*) e di serotonina cerebrale (*colonne in grigio*) nei ratti. In un'ora, una dose pari alla metà di quella di una normale razione alimentare quotidiana (*estrema destra*) fece aumentare il livello di triptofano di nove volte e quello della serotonina del 40 per cento rispetto ai livelli riscontrati in un ratto che non aveva subito l'iniezione (*estrema sinistra*). Perfino un decimo di questa dose provocò aumenti significativi nei livelli del precursore e del mediatore.

Un risultato imprevisto si ebbe in seguito al tentativo di abbassare i livelli di triptofano e di serotonina cerebrale nei ratti. In generale l'iniezione di insulina fa diminuire la quantità di amminoacidi che di norma si trovano nel sangue, così come la quantità di zucchero (*curva tratteggiata in nero*). L'iniezione provocò invece l'aumento dei livelli di triptofano nel sangue e nel cervello (*curve continue in nero e in colore*) così come il livello di serotonina cerebrale (*curva in colore tratteggiata*).

tofano ematico aumentavano ma né il triptofano cerebrale né la serotonina cerebrale subivano alcun incremento.

Abbiamo presto individuato che cosa vi era di sbagliato nella nostra previsione. Come già due gruppi indipendenti di ricercatori avevano dimostrato, gli amminoacidi sono trasportati dal sangue al cervello da un sistema di trasporto che comprende molti sottosistemi separati. Per esempio, un trasportatore trasferisce solamente gli amminoacidi che sono ionizzati alla normale acidità del sangue (pH 7). Un secondo trasportatore convoglia solamente gli amminoacidi che sono ionizzati positivamente e un terzo trasporta solo quelli che sono ionizzati negativamente. Si dà il caso che il triptofano non sia ionizzato alla normale acidità del sangue e, una volta presente nel sangue in seguito all'ingestione di proteine, esso entra in competizione per il trasporto con cinque amminoacidi anch'essi neutri (leucina, isoleucina, valina, fenilalanina e tirosina). I livelli dei cinque amminoacidi in competizione nel sangue dei ratti aumentavano notevolmente quando essi erano sottoposti a una dieta contenente proteine. L'aumento nel livello del triptofano, tuttavia, era di molto inferiore poiché il triptofano è molto meno abbondante nelle proteine che non gli altri amminoacidi. Evidentemente quando la concentrazione degli amminoacidi in competizione diventava abbastanza alta, impediva, saturando il sistema di trasporto, il passaggio di triptofano nel cervello.

Questa sequenza di eventi potrebbe anche essere responsabile dell'azione svolta dall'insulina iniettata e da quella secreta sul triptofano cerebrale. Diminuendo nel sangue la concentrazione di altri amminoacidi neutri che entrano in competizione con il triptofano per il trasporto all'interno del cervello, l'insulina faciliterebbe l'aumento del livello del triptofano cerebrale. L'aumento simultaneo nel livello di triptofano ematico indotto dall'insulina contribuirebbe ulteriormente all'aumento del triptofano cerebrale.

Per verificare questa interpretazione abbiamo preparato due nuove diete da somministrare ad altri gruppi di ratti. Una razione conteneva carboidrati, grassi, e tutti gli amminoacidi che normalmente sarebbero presenti in una dieta con il 18 per cento di caseina. La seconda razione era identica tranne per il fatto che i cinque amminoacidi neutri che entrano in competizione con il triptofano erano stati tolti. Abbiamo trovato che il livello di triptofano nel

Il paradosso dell'insulina è stato studiato in maniera più approfondita sottoponendo ratti affamati a una dieta ricca in carboidrati affinché secernessero spontaneamente insulina. Dai livelli normali (*colonna zero in ogni gruppo*) il triptofano ematico (*a sinistra*), il triptofano cerebrale (*al centro*) e la serotonina cerebrale (*a destra*) aumentarono sostanzialmente nelle prime due ore successive al momento in cui si era permesso ai ratti di accedere al cibo che avrebbe indotto la secrezione di insulina.

sangue aumentava nei gruppi di ratti nutriti con l'una o con l'altra razione. Solamente nei ratti nutriti con la razione non contenente gli amminoacidi competitori del triptofano, tuttavia, i livelli di triptofano cerebrale e di serotonina aumentavano sostanzialmente (*si veda la figura a pagina 74, in alto*). Come prova ulteriore abbiamo nutrito gruppi di ratti con una dieta mancante di due amminoacidi diversi dai precedenti, l'aspartato e la glutammina, la cui carica fa sí che essi non entrino in competizione con il triptofano. Il livello di triptofano ematico aumentava ma né il triptofano cerebrale né la serotonina cerebrale subivano alcun incremento. È chiaro che il trasporto del triptofano dal sangue al cervello è accelerato quando questo trasporto non viene ostacolato in maniera significativa da competitori del triptofano stesso.

Abbiamo riassunto le nostre conclusioni come segue. La concentrazione di triptofano nel cervello (e la conseguente sintesi di serotonina) riflette il rapporto tra il triptofano e gli altri amminoacidi neutri nel sangue piú accuratamente di quanto non faccia la quantità di triptofano ematico da sola. Una dieta ricca in carboidrati, stimolando la secrezione di insulina, tenderà ad aumentare questo rapporto; una dieta ricca in proteine, aumentando il livello di amminoacidi in competizione col triptofano nel sangue, tenderà ad abbassare questo rapporto. Abbiamo recentemente verificato questa conclusione sottoponendo un gruppo di ratti a una dieta contenente un'altissima concentrazione di proteine: 40 per cento di caseina, circa la quantità di proteine di una bistecca. La velocità della sintesi di serotonina nel cervello dei ratti effettivamente diminuí.

Tuttavia, il paradosso riguardante l'insulina non era completamente spiegato. Le nostre misurazioni mostrarono che nei ratti l'insulina agiva nel sangue non solo elevando il rapporto tra il triptofano e gli altri amminoacidi ma facendo aumentare anche la quantità di triptofano presente. Come poteva succedere? La ricerca di una risposta a questo problema, dimostrò che vi era coinvolta una componente dietetica che noi avevamo piú o meno ignorato: il grasso.

Come numerose altre sostanze che circolano nel sangue (bilirubina, alcuni ormoni e acidi grassi liberi), il triptofano ha tendenza a legarsi all'albumina, una proteina normalmente presente nel siero sanguigno. Circa l'80 per cento di tutte le molecole di triptofano nel sangue sono legate in questo modo; il fenomeno fu dimostrato per la prima volta nel 1958 da Rapier H. McMenamy e da John L. Oncley alla Harvard Medical School. Ray W. Fuller dei laboratori di ricerca della Eli Lilly & Company, che ha seguito il nostro lavoro e ci ha dato molti suggerimenti utili, ha messo in evidenza nel 1971 che alcune delle nostre osservazioni potevano dipendere da cambiamenti nel rapporto tra triptofano legato all'albumina e triptofano libero nel sangue dei soggetti.

Abbiamo discusso questa possibilità con il nostro collega Hamish Munro e abbiamo trascorso l'anno successivo lavorando con Bertha Madras e Deborah Lipsett alla ricerca del modo piú efficace di separare il triptofano libero nel siero da quello legato all'albumina. Abbiamo poi iniziato un esperimento con dei volontari allo scopo di verificare il suggerimento di Fuller. Dopo il digiuno di una notte, ogni volontario ingerí 75 grammi di glucosio; venivano poi prelevati campioni di san-

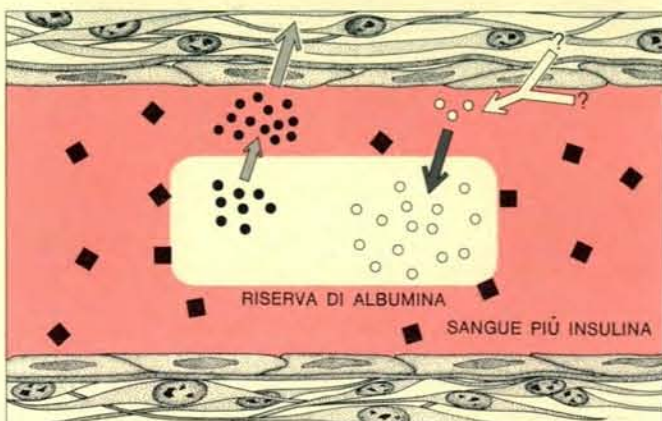
La competizione tra il triptofano e altri cinque amminoacidi per il «trasporto» dal sangue al cervello è responsabile del fatto che il livello di triptofano nel cervello non cambi dopo una dieta ricca in triptofano. Due gruppi di ratti sono stati sottoposti a diete differenti; in una (*in nero*) tutti gli amminoacidi delle proteine della dieta (compreso il triptofano e i cinque amminoacidi competitori) erano presenti. I cinque amminoacidi competitori sono stati tolti dall'altra dieta (*colore scuro*) in modo da diminuirne il livello nel sangue del secondo gruppo di ratti. Con tutti e due i tipi di dieta il livello di triptofano ematico era superiore a quello di un gruppo di controllo tenuto a digiuno (*a sinistra, in colore chiaro*). I livelli di triptofano (*al centro*) e di serotonina (*a destra*) cerebrali aumentarono però molto di piú nei ratti del secondo gruppo.

gue a intervalli regolari. Come presumevamo, l'ingestione di carboidrati causava nei volontari la secrezione di insulina, e i livelli di tutti gli amminoacidi nel sangue, tranne il triptofano, diminuivano notevolmente. Sebbene la quantità totale di triptofano ematico dei volontari rimanesse la stessa, la quantità di triptofano libero nei campioni di sangue diminuí di ben il 40 per cento. La diminuzione era accompagnata da un aumento proporzionale, anche se minore, della quantità di triptofano ematico legato all'albumina. Che cosa causava questo cambiamento?

Si sa che come la concentrazione di amminoacidi nel sangue cosí la concentrazione di acidi grassi liberi viene abbassata dalla presenza di insulina nel sangue. Sembrava possibile che, se l'insulina forzava alcune molecole di acidi grassi liberi a lasciare il proprio legame con la molecola di albumina e a lasciare il sangue, l'aumento di disponibilità di albumina non legata avrebbe permesso che un numero maggiore di molecole di triptofano passasse dallo stato libero a quello legato all'albumina.

Per verificare questa congettura abbiamo misurato la concentrazione di acidi grassi liberi nel siero dei nostri volontari. La concentrazione cambiò in modo notevolmente analogo al cambiamento di concentrazione del triptofano libero. Dopo l'ingestione di glucosio tutte e due le concentrazioni diminuirono in periodi di tempo e in proporzioni simili. Quando abbiamo ripetuto l'esperimento con ratti al posto di volontari, il risultato è stato lo stesso.

Questo ci ha condotto alla conclusione che gli acidi grassi liberi nel siero del sangue sono intermediari dell'effetto dell'insulina sul triptofano circolante nel sangue. Secondo questa ipotesi il triptofano non diminuisce nel sangue in presenza di insulina come fanno gli altri amminoacidi neutri poiché l'azione dell'insulina libera simultaneamente la albumina dai suoi legami con gli acidi grassi liberi. La liberazione dell'albumina, a sua volta, permette ad altre molecole di triptofano libero di legarsi all'albumina. Man mano che la concentrazione di triptofano libero diminuisce nel sangue, essa è parzialmente ricostituita da altre molecole di triptofano che diffondono da altri tessuti, come il muscolo scheletrico. Evidentemente nei ratti l'affinità dell'albumina per il triptofano in presenza di insulina è cosí alta che la concentrazione totale di triptofano nel sangue aumenta. Negli uomini, tuttavia, l'affinità sembra minore; di conseguenza l'aumento della quantità di triptofano legato all'albumina non fa diminuire il livello di triptofano libero nel sangue al punto da causare la diffusione di altre molecole di triptofano dai tessuti nel sangue. Negli uomini, quindi, l'insulina non dà luogo a un netto aumento nel livello di triptofano ematico.

Un'ulteriore spiegazione del paradosso dell'insulina riguarda la competizione tra le molecole di triptofano e le molecole di acidi grassi liberi nel sangue per legarsi alle molecole di albumina disponibili. Circa l'80 per cento di tutto il triptofano nel sangue è legato all'albumina in questo modo; nella figura ciò è rappresentato schematicamente dalle zone denominate «riserve di albumina». Quando l'insulina entra nel sangue (*a destra, rettangoli neri*), gli acidi grassi legati si riducono e molti di essi escono dal sangue (*frecce in grigio*). Ciò permette a un numero maggiore di molecole di triptofano libero di legarsi alle molecole di albumina liberate (*freccia in nero*); il passaggio allo stato di molecole di riserva insensibili all'insulina consente al livello totale di triptofano ematico di rimanere grosso modo in equilibrio. Nel ratto, altre molecole di triptofano migrano dai globuli rossi o dai tessuti adiacenti (*freccia doppia a destra*), aumentando cosí il livello di triptofano.

L'esistenza nel sangue di una tale riserva di triptofano legato all'albumina e insensibile all'insulina dà a questo amminoacido un vantaggio competitivo sugli altri amminoacidi neutri riguardo al trasporto dal sangue al cervello. La nostra ipotesi necessita di una verifica perlomeno per quanto riguarda un aspetto. Sarebbe certamente utile sottoporre degli individui a una dieta ricca in grassi in modo da vedere se una sovrabbondanza di acidi grassi liberi blocca nel sangue abbastanza albumina extra da limitare o sopprimere il legarsi del triptofano all'albumina e ridurre cosí la riserva di triptofano legato insensibile all'insulina.

Siamo convinti che il consumo di cibi a basso tenore proteico produca un aumento nella sintesi di serotonina nel cervello dei ratti, anche se si è dimostrato che la causa di ciò è abbastanza complessa. Gli stessi metodi di dosaggio non possono essere adottati su volontari, cosicché non abbiamo prove tangibili che lo stesso sia vero per l'uomo. Tuttavia, nutrendo i nostri volontari con sostanze senza proteine e con sostanze ricche in proteine e misurando poi la quantità di triptofano e di altri amminoacidi neutri nel loro sangue, abbiamo stabilito che gli uomini rispondono esattamente come i ratti agli stessi tipi di dieta. Sembra logico presumere che i cambiamenti nel livello di triptofano nel cervello umano si riflettano nel livello di serotonina cerebrale.

Non abbiamo ancora trovato una prova diretta che la concentrazione di questo mediatore particolare nei neuroni dei nuclei del rafe induca necessariamente dei cambiamenti nella attività funzionale dei neuroni. L'intuizione, tuttavia, ci spinge a trarre questa conclusione e a supporre che la sensibilità delle concentrazioni di serotonina cerebrale a cambiamenti indotti dalla dieta probabilmente fornisce ai mammiferi, incluso l'uomo, un meccanismo utile di transduzione.

Sembra ragionevole ipotizzare che i cambiamenti indotti dalla dieta nella concentrazione di serotonina nei neuroni dei nuclei del rafe vadano di pari passo con cambiamenti nella quantità di serotonina liberata nelle sinapsi ogni volta che i neuroni trasmettono segnali o sono forzati a farlo da un intervento chimico. Se è cosí, allora i neuroni serotoninergici funzionerebbero come sensori, che convertono un segnale che circola nel sangue in un ordine nel linguaggio del cervello. Una dieta regolare ricca in proteine avrebbe come conseguenza una liberazione minima di serotonina da parte dei neuroni dei nuclei del rafe, mentre una dieta ricca in carboidrati, anche se somministrata per un breve periodo, causerebbe una liberazione massima del mediatore.

Non sembra infondato pensare che il cervello possa fare buon uso di queste risposte al rapporto proteine-carboidrati della dieta e forse ad altri stimoli nutrizionali e ormonali che sono percepiti nello stesso modo. I neuroni serotoninergici possono portare al resto del cervello dati importanti non solo riguardo alle conseguenze metaboliche di cambiamenti a breve e a lungo termine nella nutrizione ma anche riguardo ai processi di crescita e di sviluppo e riguardo allo stress.

Si sa che i neuroni serotoninergici di alcuni mammiferi sono associati a numerosi meccanismi neuroendocrini. Per esempio, i gatti privati di serotonina cerebrale diventano insonni, e la riduzione nel livello di serotonina cerebrale nei ratti maschi e femmine suscita una esagerata attività sessuale; un aumento nel livello della serotonina cerebrale diminuisce, invece, la sensibilità del ratto al dolore. Anche la secrezione di numerosi ormoni pituitari è in relazione con il livello di serotonina cerebrale. Un recente lavoro che abbiamo compiuto con il nostro collega Loy Lytle suggerisce che i cambiamenti nei livelli di serotonina cerebrale nei ratti possono influenzare sia l'attività motoria sia il consumo di cibo. Quest'ultimo risultato suggerisce un tipo di circolo chiuso, dove il consumo di cibo influirebbe sulla biochimica del cervello e la biochimica del cervello a sua volta influirebbe sul consumo di cibo.

È probabile che questo particolare circolo chiuso sia una spirale vista di fronte, o anche una curva che non conduce da nessuna parte. Tuttavia resta il fatto che, almeno nel mammifero nostro parente, ossia il ratto di laboratorio, la dieta controlla in effetti la sintesi di un mediatore significativo. Siamo riluttanti a liquidare questa scoperta come una pura coincidenza.

Il controventamento dei fabbricati

L'evoluzione dei grattacieli ha suggerito complessi sistemi di irrigidimento contro la spinta del vento. Per i fabbricati piú alti è stata rimessa in uso una tecnica di oltre 100 anni fa

di Carl W. Condit

Un fabbricato è in generale sottoposto a tre tipi di forze. La prima è il cosiddetto « carico morto» cioè il peso proprio della struttura e del suo contenuto stabile; la seconda è il « carico vivo », ovvero il peso degli occupanti e delle parti mobili, come per esempio gli ascensori. Si tratta in entrambi i casi di forze gravitazionali che devono essere in definitiva trasmesse al terreno, lungo linee verticali, tramite le fondazioni.

La terza forza è prodotta dal movimento dell'aria e agisce in direzione perpendicolare alle pareti dell'edificio, con effetti differenti dalle prime due. Infatti i carichi gravitazionali sollecitano le strutture portanti verticali (colonne e pilastri) prevalentemente a compressione; l'azione del vento induce invece nell'ossatura della costruzione sollecitazioni complesse di flessione, torsione, taglio e rovesciamento.

Ai tempi in cui gli edifici venivano costruiti in muratura massiccia gli effetti del vento potevano essere praticamente trascurati nel progetto della struttura. Una costruzione in grado di reggere l'enorme peso proprio di spesse pareti in pietra è intrinsecamente indifferente alla spinta del vento. Quando però si cominciarono a erigere fabbricati di altezza eccezionale sostenuti da una leggera intelaiatura in ferro o in acciaio, il problema di assicurare una adeguata resistenza al vento si dimostrò di difficile soluzione. I metodi costruttivi seguiti in precedenza potevano fornire solo scarse indicazioni: infatti molti degli artifici e delle tecniche adottati dai costruttori di grattacieli non sono derivati dalla precedente esperienza edilizia, ma dai metodi ideati per la costruzione di ponti a grande luce libera.

L'integrità strutturale di un fabbricato dipende da due proprietà. Una struttura, considerata come un tutto, si definisce stabile se rimane immobile rispetto alla superficie di appoggio sotto l'azione dei carichi previsti in sede di progetto; se invece i vari elementi di cui è costituita non subiscono spostamenti relativi, viene detta rigida.

Un telaio rettangolare semplice è stabile sotto l'azione di un carico verticale fisso (se i singoli elementi hanno una resistenza sufficiente), ma non è invece rigido nei confronti di una forza orizzontale o anche di una forza con una componente orizzontale. Sottoposto all'azione di un carico di questo genere, il rettangolo si distorce, diventando un parallelogrammo e può al limite essere perfino del tutto appiattito (*si veda l'illustrazione nella pagina successiva in alto*).

Questa debolezza può venire eliminata collocando un controvento diagonale, che trasforma il telaio in una struttura rigida composta da due triangoli (il triangolo è una figura rigida, che non può essere deformata senza distorcere uno o piú dei lati). Con l'aggiunta del controvento il telaio acquisisce entrambe le proprietà di stabilità e di rigidità: se gli elementi che lo compongono sono sufficientemente resistenti e se i giunti sono abbastanza forti, esso può sopportare sollecitazioni sia verticali sia orizzontali. L'unico difetto è ora la presenza dell'elemento diagonale che attraversa l'intero telaio e che può rendere inutilizzabile lo spazio interno. Occorre pertanto trovare altri metodi per irrigidire le strutture, metodi che costituiscono una gran parte della moderna scienza delle costruzioni.

Il carpentiere medievale si era reso ben conto della necessità di irrigidire i telai in legno contro le spinte del vento. Una copertura a falde sorretta da capriate (o incavallature) triangolari è intrinsecamente rigida, ma la catena della capriata e i piedritti che la sostengono formano un rettangolo che non lo è. Non è possibile collocare un elemento lungo la diagonale del rettangolo, in quanto ostruirebbe l'intera luce (la luce è la distanza tra ogni coppia di piedritti o montanti adiacenti di ciascuna unità dell'intelaiatura). La soluzione piú semplice del problema fu l'applicazione di un piccolo elemento diagonale fra la catena e il piedritto, denominato controvento a ginocchio, che irrigidisce il collegamento occupando spazio solo vicino al tetto. Un artificio di maggiore robustezza ed eleganza è stato il controvento curvo fissato con giunti a tenone e mortasa alla capriata e al montante. Tali controventi sono stati detti « a ginocchio di nave » perché derivati dagli elementi curvi chiamati ginocchi che collegano nelle navi a vela i bagli del ponte con le coste dello scafo. Controventi analoghi sono stati disposti anche in direzione parallela all'asse longitudinale dell'edificio (*si veda l'illustrazione in basso nella pagina successiva*).

Le spinte del vento sui grossi fabbricati in muratura non hanno costituito un problema fino al XII secolo. In precedenza erano state usate grandi quantità di pietra per sostenere il peso proprio della struttura e sulle pareti spesse e pesanti il vento rappresenta una sollecitazione trascurabile. Tuttavia i costruttori del periodo gotico trasformarono la pesante chiesa romanica in una struttura estremamente alleggerita costituita da costoloni, pilastri e contrafforti che formavano quasi un'intelaiatura in pietra. Per questa struttura relativamente debole la resistenza al vento assume una notevole importanza (si veda *L'analisi strutturale delle cattedrali gotiche* di Robert Marx in « Le Scienze » n. 54, febbraio 1973).

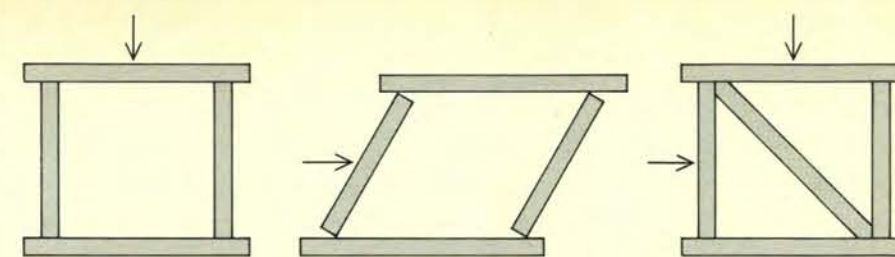
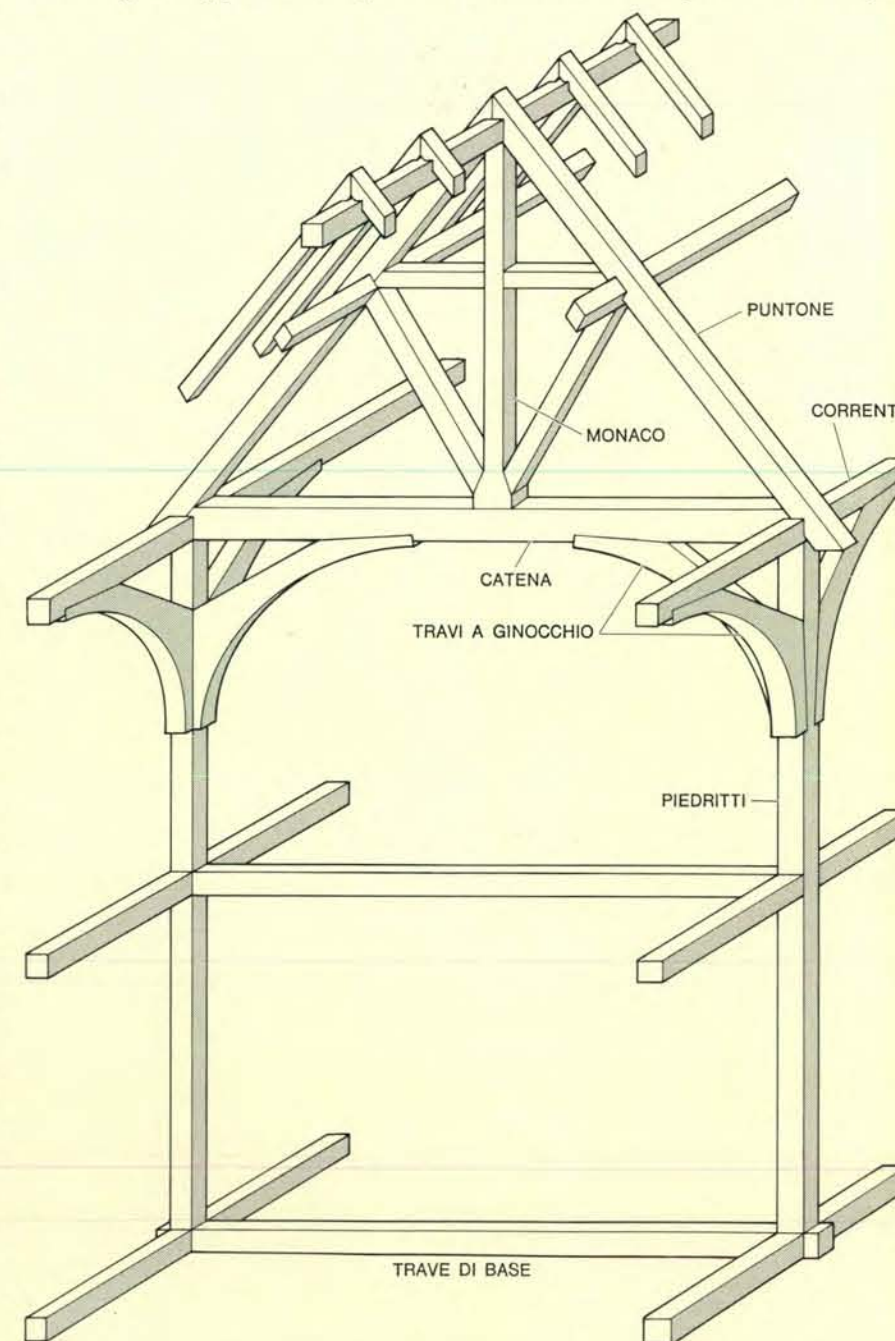
Durante il Rinascimento gli architetti tornarono a costruire con la classica tecnica delle pareti in muratura piena, già adoperata in precedenza negli edifici di dimensioni monumentali, e i problemi derivanti dalla resistenza al vento tornarono a essere trascurabili. Le soluzioni escogitate dai costruttori medievali per i piú piccoli edifici in legno erano soddisfacenti e furono impiegate usualmente ben oltre l'inizio del XIX secolo.

Con la rivoluzione industriale del 1700 sorse uno spirito piú innovatore, ispirato da un insieme di fattori correlati, che promosse l'introduzione del ferro come materiale strutturale primario. Il nuovo impiego del ferro portò come conseguenza due nuovi indirizzi nella scienza delle costruzioni: le accresciute dimensioni delle strutture a telaio con la contemporanea riduzione delle sezioni degli elementi portanti, l'uno, e la trasformazione dell'edilizia da mestiere pragmatico e empirico in una tecnologia affermata sulla base di una scienza teorica e sperimentale, l'altro.

All'inizio i cambiamenti furono lenti. Cosí, per esempio, ancora nel 1850 l'aspetto di fabbriche e magazzini a piú piani con struttura portante in ferro non differiva in modo sensibile da quello dei rispettivi prototipi con struttura portante in legno.

Le prime costruzioni con elementi strutturali principali in ferro furono i cotonifici in cui vennero alloggiati i nuovi macchinari tessili azionati a motore. I progettisti seguirono la tradizionale tecnica di supporto della struttura esterna con muratura di mattoni pieni, sostituendo però il ferro al legno nei piedritti e nelle travi interne. Per l'elevato peso della muratura, non si ritenevano necessarie le strutture interne di irrigidimento, opinione che permase fino all'ultimo decennio del XIX secolo.

È stata l'introduzione del telaio semplice in ferro appoggiato al suolo a far comprendere ai costruttori la necessità del controventamento, trattandosi di strutture prive dell'involucro protettivo di muratura. La capriata in ferro venne impiegata per la prima volta nel 1810 e nel 1820 i costruttori cominciarono a sostituire le pareti in muratura con colonne in ferro. Uno dei primi esempi di questa nuova struttura è stato un piccolo mercato parigino chiamato la Madeleine, costruito nel 1824, la cui intelaiatura e le cui capriate erano del tutto prive di controventamento. Un'opera piú progredita è lo Hungerford Fish Market di Londra, progettato dall'architetto Charles Fowler ed edificato nel 1835. L'intelaiatura del tetto è sorretta da colonne cilindriche in ghisa e l'irrigidimento è assicurato da controventi in ferro a ginocchio di nave, sporgenti da collari fissati alle colonne.

I rapidi progressi conseguiti nelle costruzioni in ferro durante la restante parte del 1800 sono dovuti all'applicazione della scienza ai problemi di progetto delle strutture, ai metodi piú

Un telaio rettangolare serve a dimostrare il significato delle proprietà di stabilità e di rigidità. Il telaio è stabile sotto l'azione di un carico verticale (*figura a sinistra*), rimanendo immobile rispetto alla superficie di appoggio; se invece esso viene sottoposto a una forza orizzontale non è piú rigido (*figura al centro*) e i suoi elementi subiscono spostamenti relativi. La struttura si può peraltro rendere rigida collocando un controvento diagonale (*figura a destra*), che suddivide l'intero rettangolo in due triangoli.

Il telaio in legno della casa medievale è stato una delle prime strutture che hanno richiesto l'irrigidimento contro le spinte del vento. La capriata formata dai puntoni e dalla catena è un triangolo ed è intrinsecamente rigida, ma la catena forma con il resto del telaio un rettangolo che invece non lo è (*vedasi la figura in alto*). Poiché un elemento disposto lungo la diagonale del rettangolo ostruirebbe la campata, sono stati impiegati degli elementi diagonali piú piccoli, denominati controventi o travi « a ginocchio », montati fra la catena e i piedritti e fra questi ultimi e i correnti.

accurati per l'analisi delle sollecitazioni nelle strutture a telaio e ai laminatoi capaci di produrre profili in ferro aventi altezza fino a 180 millimetri. Il lavoro che ha maggiormente anticipato il futuro è stato il Palazzo di Cristallo, eretto nel 1851 in occasione dell'esposizione di Londra dall'architetto sir Joseph Paxton, orticoltore e costruttore di serre a telaio in ferro.

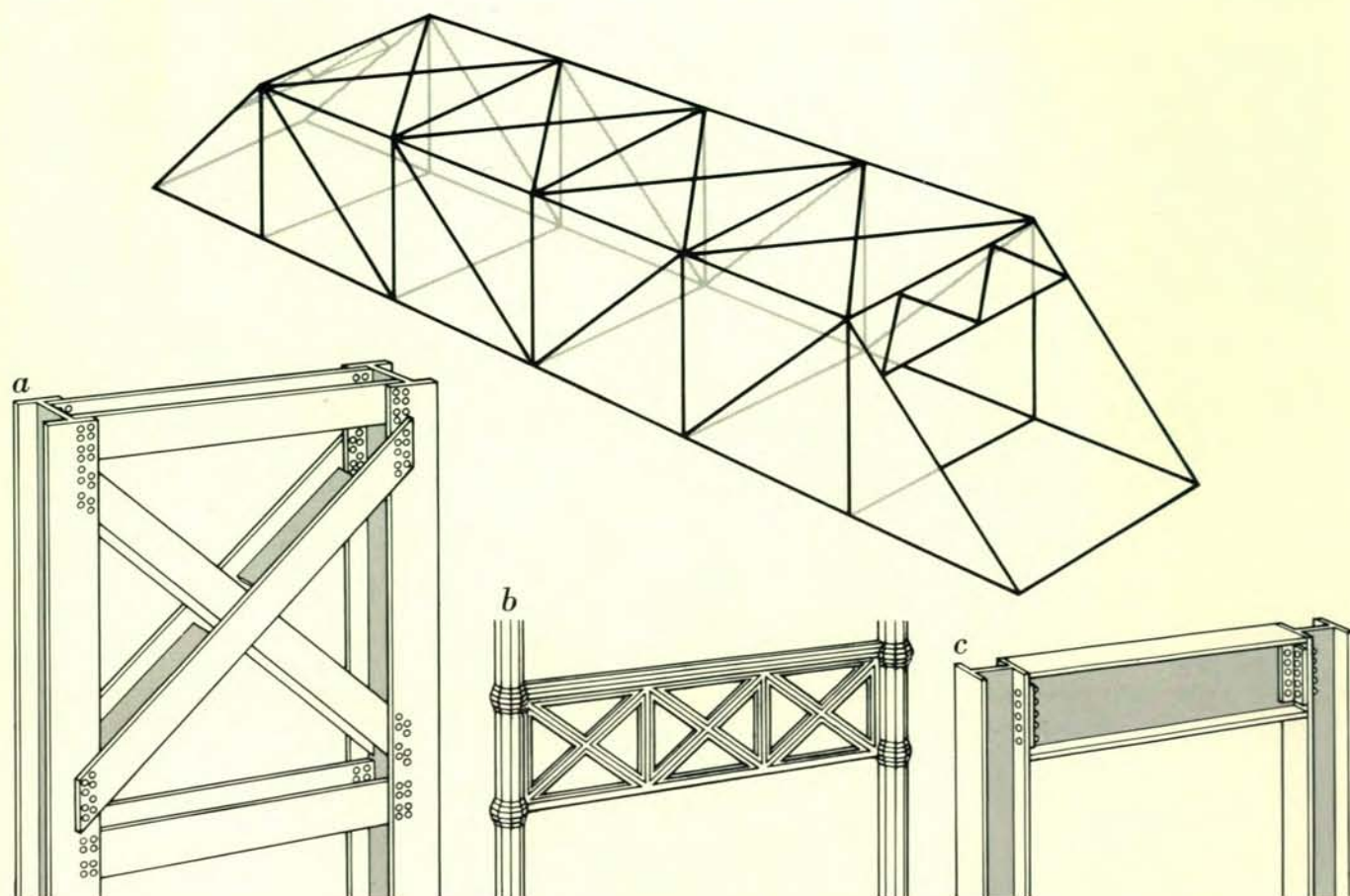
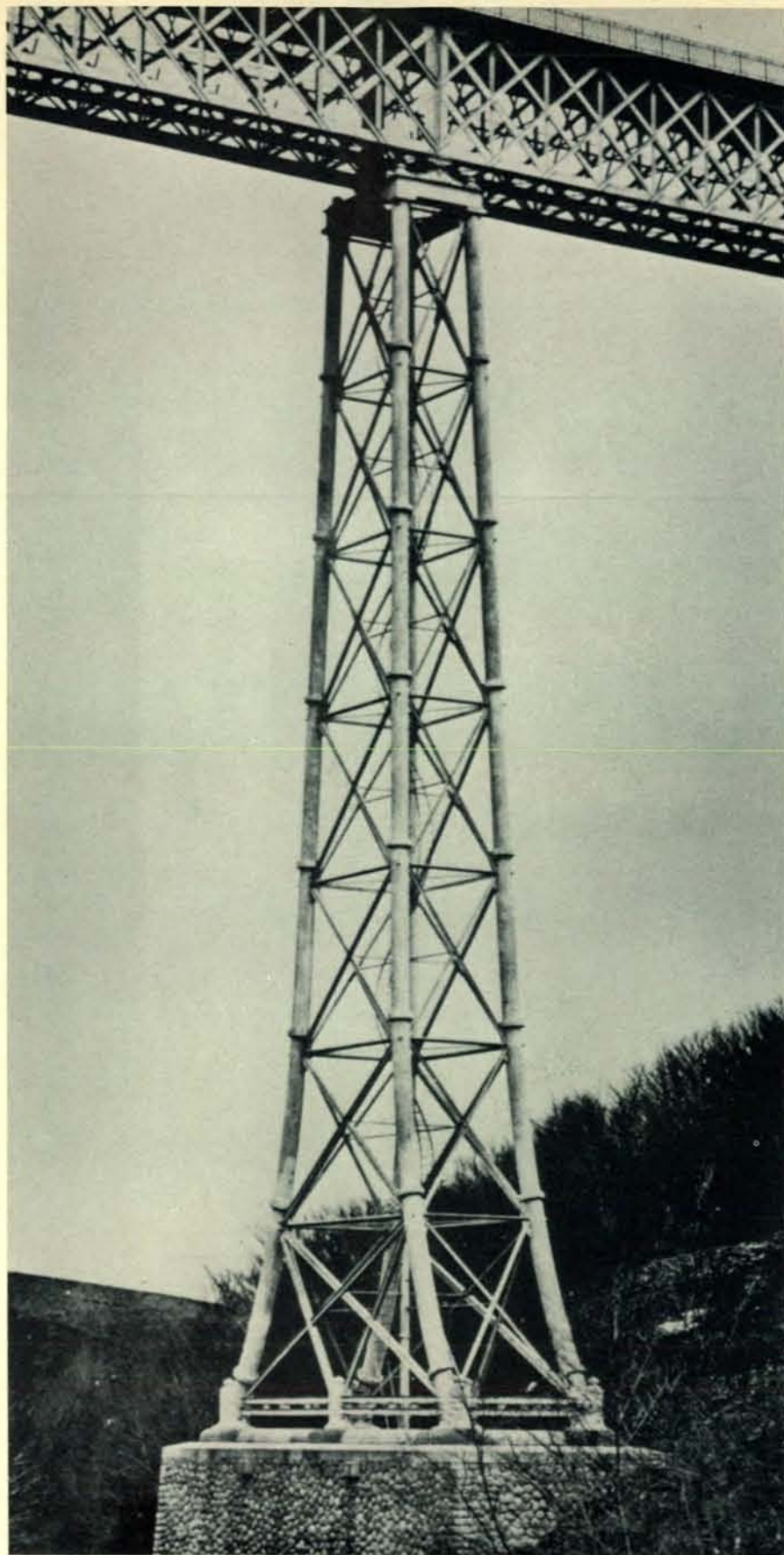
Il Palazzo di Cristallo è stato il primo edificio di grandi dimensioni realizzato con intelaiatura in ferro: misurava in pianta 144,5 per 563,3 metri ed è stato anche il primo fabbricato con tamponature esterne in vetro, il primo costruito con unità strutturali prefabbricate e il primo nel quale leggeri telai sono stati irrigiditi contro la spinta del vento con una tecnica nota in seguito con il nome di controventatura a portale.

Nel Palazzo di Cristallo sono stati impiegati due tipi di controventatura. Il primo tipo era costituito da membrature diagonali disposte in croce per formare una serie di X, realizzate con tondini di ferro battuto e connesse agli elementi portanti principali della volta centrale. Il secondo era applicato nei punti ove le travi reticolari in ghisa che sorreggevano le gallerie incontravano le colonne ottagonali: dei collari a flangia fissati alle colonne erano chiodati alle facce superiore e inferiore delle travature, eliminando la necessità di elementi diagonali per la presenza dei due punti di fissaggio. Entrambe le tecniche erano state derivate dai metodi impiegati nella costruzione di ponti a trave reticolare, che sono tradizionalmente controventati con membrature diagonali incrociate disposte tanto superiormente che inferiormente. Il secondo tipo di controvento usato nel Palazzo di Cristallo si trova nei telai di estremità dei ponti citati, denominati telai porta (o di ingresso) la cui membratura di irrigidimento viene chiamata controventatura a portale (*si veda l'illustrazione in questa pagina*).

Il principio su cui si basa questa tecnica è antico almeno quanto la carpenteria primitiva: due tavole unite con un solo chiodo ruotano con facilità intorno al chiodo stesso, ma se si usano due chiodi il giunto diventa rigido. Nel Palazzo di Cristallo il metodo è stato probabilmente adottato per motivi estetici poiché i controventi a ginocchio sarebbero sembrati privi di grazia nell'elegante traliccio.

La controventatura a portale nella forma moderna apparve pochi anni dopo. Il primo esempio si trova nel Royal Navy Boat Store a Sheerness in Inghilterra, che fu progettato da Godfrey T. Greene e costruito fra il 1858 e il 1860. Alte travi in ferro battuto erano chiodate a colonne in ghisa per tutta la loro altezza. Ulteriori perfezionamenti della tecnica sono stati introdotti nei grattacieli che cominciarono a essere costruiti a New York e a Chicago alla fine del secolo.

Nello stesso periodo parecchi costruttori introdussero altre tecniche fondamentali nella costruzione dei moderni edifici a grande sviluppo verticale. Sebbene sembrasse che tali progressi non fossero in relazione fra loro, cominciò a manifestarsi un nuovo esempio di evoluzione tecnologica.

Numerose sono le novità applicate per la prima volta nella costruzione dei fari. La forma di massima stabilità al-

I disegni illustrano la derivazione dai ponti a trave reticolare di due tecniche di controventamento. Il controventamento con membrature diagonali incrociate, disposte in genere nei pannelli superiori e inferiori del ponte, è stato adottato direttamente per i fabbricati dove l'ostruzione dei volumi interni non è inaccettabile (*a*). La controventatura a portale deriva dai telai di ingresso, o telai porta, del ponte. Nei primi esempi, quale il Palazzo di Cristallo, una trave reticolare era fissata alla colonna in due punti (*b*). Negli edifici costruiti successivamente le travi vennero chiodate ai montanti per tutta la loro altezza (*c*).

La svasatura dei montanti riduce le sollecitazioni nella struttura portante, o torre di sostegno, del viadotto Bellon della linea ferroviaria Gannat-Commentry in Francia. Le sollecitazioni di taglio e di flessione raggiungono i massimi valori alla base del traliccio; la curvatura dei montanti tende a trasformare il taglio in compressione, cioè in un tipo di sollecitazione alla quale gli elementi in ghisa resistono meglio. Consulente per il progetto del viadotto, che fu costruito fra il 1868 e il 1871, fu Gustave Eiffel.

la spinta del vento è per queste strutture quella a « tronco d'albero », cioè di un cilindro rastremato con la parte larga alla base. È la forma usata da John Smeaton già nel 1759 per la torre in muratura del faro Eddystone in Inghilterra, adattata poi alle costruzioni in ferro da parte degli ingegneri francesi che l'hanno perfezionata per il faro di Roches-Douvres sulla Manica, eretto negli anni 1860-1870.

Il faro di Roches-Douvres, alto 48 metri, era rivestito con piastre di ferro, applicate su un traliccio in ferro battuto realizzato con colonne, travi periferiche e controventi diagonali. La resistenza al vento era ottenuta con tre tipi di accorgimenti: le ormai familiari membrature diagonali nei primi tre ripiani in basso, ove i momenti flettenti indotti dal vento raggiungono i massimi valori; le piastre nodali di rinforzo nelle giunzioni chiodate fra le colonne, le travi e gli elementi diagonali e infine la larga base dalla forma a tronco d'albero.

Nello stesso decennio gli ingegneri progettisti della linea ferroviaria Gannat-Commentry in Francia svilupparono dei tralicci a tronco d'albero per le strutture portanti in ferro, o torri di sostegno, dei numerosi ponti nel tratto montano del percorso. Fra queste strutture, le piú innovatrici furono progettate per il viadotto Bellon, costruito fra il 1868 e il 1871 (*si veda l'illustrazione a sinistra*).

Le torri a traliccio sono realizzate con quattro montanti tubolari in ghisa legati a intervalli da traverse orizzontali e controventati da tondini diagonali incrociati in ciascuno degli specchi verticali e orizzontali delimitati dai montanti e dalle traverse. La caratteristica singolare della struttura è la svasatura alla base, esattamente calcolata, imposta ai montanti che non solo assicura una maggiore larghezza nella zona ove i momenti flettenti sono maggiori, ma consente anche di trasformare nei robusti tubi di ghisa, in corrispondenza dell'incastro nel basamento, le sollecitazioni di taglio in sollecitazioni di compressione. I ponti della ferrovia Gannat-Commentry sono stati progettati dagli ingegneri Delom, Geoffroy, Nordling e Thirion. Consulente per il progetto fu un ben piú illustre ingegnere: Gustave Eiffel, l'uomo che nel 1870 era la massima autorità europea per il calcolo della stabilità aerodinamica delle strutture di grande altezza.

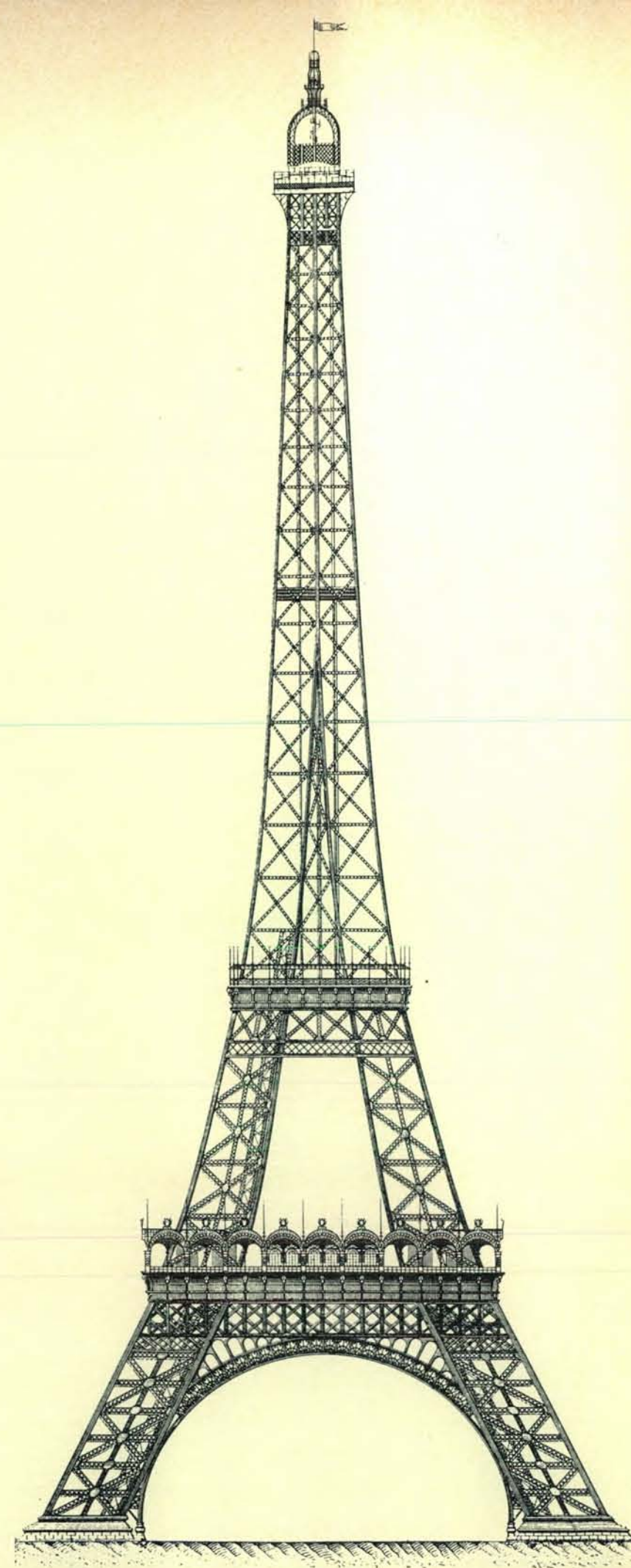
La celebre Torre Eiffel, progettata per l'Esposizione di Parigi del 1889, è l'esempio piú eminente di un'opera eretta seguendo rigorosamente le norme della scienza delle costruzioni e dell'aerodinamica. I principi dell'aerodinamica applicati nella sua realizzazione guidano tuttora il progetto dei grattacieli a traliccio in acciaio.

L'immensa costruzione in ferro, un *tour de force* eccezionale che non è niente altro che un monumento a se stessa, venne eretta fra il 1887 e il 1889 su progetto di Eiffel, assistito dagli ingegneri Emile Nouguier e Maurice Koechlin e dall'architetto Shephen Sauvestre. La torre doveva essere alta 300 metri e divisa da tre ripiani, il primo per poter ospitare un ristorante e gli altri due per fungere da belvedere.

Negli anni 1880-1890 la costruzione dei ponti di grande luce era diventata una scienza prevedibile e quindi il calcolo delle forze gravitazionali era basato su un'esauriente casistica. Eiffel sapeva che il vento avrebbe costituito il fattore decisivo del progetto e che la resistenza a esso avrebbe determinato la stabilità o la caduta della torre.

Egli prese fin dall'inizio la decisione di trattare l'intera struttura come una mensola verticale (una mensola è una trave o una struttura reticolare vincolata a uno solo degli estremi). La mensola è costituita da due parti: una è la torre vera e propria compresa fra il secondo ripiano e il piano di calpestio delle attuali stazioni meteorologica e televisiva poste sulla cima; l'altra è distinta dalla prima anche in modo visivo ed è formata dal traliccio di base, chiamato da Eiffel il *montant* (*si veda l'illustrazione a destra*). La struttura di transizione è un traliccio periferico che funge da enorme cravatta e che è situato immediatamente al di sotto del secondo ripiano.

La torre è quello che oggi sarebbe definito un « telaio spaziale », cioè un traliccio tridimensionale che racchiude uno spazio definito e i cui elementi sono tutti connessi fra loro e si comportano come una struttura unica. La torre è delimitata da un inviluppo rastremato curvilineo che prolunga con eleganza la pronunciata curvatura della base. Le singole strutture portanti vuoi della torre vuoi del *montant* derivano direttamente dalle torri di sostegno dei

Antenata dei moderni grattacieli, la Torre Eiffel è stata calcolata dal progettista come una mensola verticale, cioè come una struttura monolitica incastrata a un estremo e libera di spostarsi all'altro. La sua forma a guglia curva e rastremata riduce l'area esposta al vento in prossimità della cima e converte le sollecitazioni di flessione e taglio in sollecitazioni di compressione alla base. La torre, costruita per l'Esposizione di Parigi del 1889, è alta 300 metri. Il disegno è tratto dallo studio di Eiffel *La Tour de Trois Cents Mètres*.

ponti: ciascuna è costituita da quattro montanti collegati da puntoni orizzontali e da controventi diagonali incrociati.

Eiffel aveva posto a base del calcolo due ipotesi di sollecitazione prodotta dal vento; una costituita da un carico uniforme dalla cima alla base di 300 chilogrammi per metro quadrato, l'altra da un carico variabile linearmente fra 400 chilogrammi per metro quadrato alla cima e 200 chilogrammi per metro quadrato al suolo.

Il carico equivalente nell'ipotesi di vento uniforme (cioè il vettore che rappresenta la risultante delle forze agenti per effetto del vento) risultava applicato circa a metà altezza della costruzione, con un momento flettente massimo alla base della torre di 50 923 218 chilogrammetri e uno sforzo di taglio massimo di 66 960 chilogrammi. Nella ipotesi di vento variabile con linearità il momento flettente raggiungeva il valore di 59 745 250 chilogrammetri e lo sforzo di taglio i 73 050 chilogrammi. Se i supporti della torre fossero stati perpendicolari al terreno, il momento flettente al suolo sarebbe stato di 225 milioni di chilogrammetri. Eiffel ridusse in modo drastico questa enorme forza di rovesciamento, trasferendola al vincolo tramite i quattro tralicci portanti fortemente incurvati che compongono la struttura di base.

L'insolita forma del *montant*, l'idea più singolare del progetto, venne scelta dopo un'analisi rigorosa. Essa si ritrova nella sagoma svasata del faro di Roches-Douvres e nelle torri di sostegno del viadotto ferroviario Bellon. Eiffel e i suoi assistenti dettero alla configurazione una base scientifica.

La curva di ciascuna delle quattro strutture portanti segue, per quanto è consentito a una costruzione rigida metallica, la curva che rappresenta la diminuzione del momento flettente con l'altezza. La curvatura dei tralicci si approssima a un segmento di parabola, ovvero alla medesima curva che si ottiene tracciando il reciproco del momento flettente in funzione dell'altezza. L'asse dei tralicci corrisponde con buona approssimazione alla linea di pressione lungo la quale viene trasmesso il peso proprio della costruzione (*si veda l'illustrazione in basso*). La conseguenza pratica della sagoma scelta è che gli enormi momenti flettenti e gli elevatissimi sforzi di taglio nel *montant* vengono progressivamente trasformati in una sollecitazione di compressione, ossia nel tipo di sollecitazione ideale per le strutture portanti rigide e cave. Infine la forma prescelta riduce le oscillazioni, sebbene queste restino ancora considerevoli nel tratto sottile della torre.

L'andamento dei momenti flettenti nella struttura portante di base, o *montant*, della Torre Eiffel corrisponde alla forma della torre stessa. I momenti flettenti tracciati in funzione dell'altezza seguono un segmento di parabola; i quattro tralicci che compongono il *montant* approssimano la stessa curva. Il momento flettente nell'ipotesi di una spinta uniforme del vento pari a 300 chilogrammi per metro quadrato (*curva nera continua*) e quello nell'ipotesi di una spinta del vento variabile fra 200 chilogrammi per metro quadrato alla base e 400 chilogrammi per metro quadrato al pinnacolo (*curva nera tratteggiata*) si discostano solo leggermente dalla linea di mezzeria del traliccio (*curva tratteggiata in colore*). Anche la linea di pressione (*curva in grigio*) lungo la quale viene trasmesso il peso proprio della costruzione segue la sagoma del traliccio.

La torre di Parigi era stata preceduta dalla statua della Libertà, eretta fra il 1883 e il 1886 sull'isola che a quel tempo era chiamata di Bedloe, nel porto di New York. La statua è opera dello scultore Frédéric Auguste Bartholdi mentre il traliccio interno di supporto fu progettato da Eiffel. Costituita da un involucro cavo in rame che si innalza per 57 metri e 37 centimetri al di sopra del piedistallo in muratura, più alta di qualsiasi edificio allora esistente e superata solo da alcuni piloni di ponti, la statua ha posto dei problemi particolari di controventamento. La struttura centrale, costituita da quattro montanti legati da puntoni orizzontali e irrigidita da diagonali incrociate in ciascuna delle quattro facce, sostiene un traliccio sul quale sono fissati i pannelli in rame. La disposizione delle diagonali e la struttura a traliccio dimostrano che Eiffel aveva considerato la figura principale come una mensola verticale e il braccio sollevato come una mensola aggiunta eccentricamente alla prima.

Fu necessario attendere il 1960 perché il modo di affrontare il progetto delle strutture a forte sviluppo verticale usato da Eiffel fosse applicato alla costruzione dei grattacieli. E venne adottato allora perché gli edifici a torre di grande altezza si dimostrano più vulnerabili al vento di qualsiasi altra costruzione precedente, sia per il grande valore del rapporto fra l'altezza e la sezione orizzontale sia per la mancanza delle tradizionali pareti portanti in muratura.

Fra i primi grattacieli vi sono l'Havemeyer Building a New York e il Monadnock e il Manhattan Building a Chicago, nei quali la resistenza al vento è affidata a diagonali incrociate e a controventature a portale; i primi due hanno anche spesse pareti in muratura alla base.

Corydon T. Purdy ha sviluppato un tipo di controventatura più elaborato e costoso, conosciuto come arco a portale, che venne impiegato negli alti, stretti e sottili grattacieli frequenti a Chicago nell'ultimo decennio del XIX secolo. Egli progettò per l'Old Colony Building, edificato fra il 1894 e il 1895, un complicato sistema, derivato ancora una volta dal telaio porta dei ponti a trave reticolare, nel quale delle piastre di acciaio erano chiodate fra loro in modo da formare un arco pieno, chiodato a sua volta per tutta l'altezza ai pilastri e, per la larghezza della luce, alle facce inferiori delle travi di solaio (*si veda l'illustrazione in questa pagina*).

L'arco a portale, derivato dalla controventatura a portale, consentì ai costruttori nell'ultimo decennio del XIX secolo, di erigere fabbricati alti e stretti privi di pesanti pareti portanti in muratura. L'Old Colony Building di Chicago, costruito nel 1893-94 era irrigidito con archi a portale in ciascuno dei suoi diciassette piani (*vista in sezione a sinistra*). Le luci centrali erano libere, mentre le serie di archi sovrapposti nelle luci laterali formavano delle strutture trasversali rigide in grado di sostenere la costruzione. Ogni arco era costruito con piastre piane in acciaio, chiodate alle flange delle colonne cilindriche (*dettaglio a destra*). Il sistema era costoso sia per il materiale che per la mano d'opera richiesti e inoltre ostruiva lo spazio interno.

Il sistema degli archi a portale scelto da Purdy era costoso sia per la quantità di materiale sia per la mano d'opera che richiedeva, ma la resistenza che assicurava alla struttura indusse gli ingegneri della Gunvald Aus Company ad adottarlo per il Woolworth Building di New York. Costruito fra il 1911 e il 1913, il Woolworth Building, di 231,80 metri, rimase l'edificio più alto del mondo per quasi venti anni (la torre Eiffel era però ancora più alta di circa 46 metri). Gli archi a portale irrigidiscono la struttura fino al ventottesimo piano, mentre nei piani superiori la funzione è affidata a controventi a ginocchio, chiodati nelle facce superiore e inferiore delle travi trasversali principali nei punti di incontro con i pilastri.

Il continuo aumento nell'altezza dei grattacieli di New York, dai 55 piani del Woolworth Building ai 102 dell'Empire State Building, insieme con i costi crescenti del materiale e della mano d'opera, resero ben presto economicamente proibitivo l'arco a portale. Questo, oltre a essere costoso, richiedeva anche il sacrificio di molto spazio, che avrebbe potuto essere utilmente impiegato in altro modo. Di conseguenza cominciarono a prevalere negli edifici a forte sviluppo verticale altre forme di gomiti, diagonali e controventature a portale più antiche e più leggere, anche se spesso non meno ineleganti.

Alcune caratteristiche del progetto costruttivo e architettonico dei grattacieli permisero in quel periodo di sfruttare per la controventatura alcuni semplici accorgimenti. La distanza fra i pilastri superava di rado 9 metri, rimanendo spesso al di sotto di 6 metri. Le tramezzature erano realizzate in pesanti mattoni forati e separavano i vani dal pavimento al soffitto; l'area complessiva destinata alle finestre era meno della metà della intera area delle pareti, mentre le singole finestre avevano dimensioni solo approssimativamente standard. La muratura di tamponamento fra le file verticali di finestre era rifinita con un sottile strato di pietra calcarea e i pannelli (che separavano verticalmente le aperture) erano spesso in acciaio inossidabile. Le conseguenze costruttive di queste caratteristiche erano un elevato peso proprio per unità di area coperta e una fitta schiera di pilastri su cui potevano distribuirsi le sollecitazioni di flessione e taglio indotte dal vento. Il grattacielo possedeva quindi una certa resistenza intrinseca al vento, analoga a quella della chiesa romanica, e il suo progettista si preoccupava tanto dei carichi gravitazionali quanto della resistenza al vento della costruzione.

Tuttavia, quando nel 1950 si riprese la costruzione di fabbricati commerciali, dopo il lungo intervallo dovuto alla crisi, alla seconda guerra mondiale e alle incertezze del dopoguerra, si affermò un nuovo stile architettonico per l'influenza degli architetti europei

I metodi tradizionali di controventamento pervennero al culmine dello sviluppo nell'edificio della Chase Manhattan Bank a New York City, completato nel 1961. La torre di 60 piani, alta 240 metri, è stata progettata dall'ufficio di New York della società Skidmore, Owings & Merrill e rappresenta il limite massimo di applicazione della scienza delle costruzioni per gli edifici sostenuti da un'ossatura interna di travi controventate in acciaio e racchiusi da una leggera parete «a cortina» in vetro e metallo.

Mies van der Rohe e Le Corbusier. La moda dominante impose i grattacieli con pareti interamente di vetro mentre la richiesta di superfici interne non limitate da tramezzature fisse portò all'adozione di pannelli leggeri e mobili e a luci fra i montanti di 12 o anche piú metri.

All'inizio gli ingegneri affrontarono il nuovo stile architettonico con le tecniche tradizionali, le cui caratteristiche fondamentali, come si è visto prima, possono essere fatte risalire alla metà del XIX secolo. L'opera che rappresenta la conclusione di questo sviluppo durato un intero secolo e che nel contempo chiarisce i motivi per cui era indispensabile escogitare nuove tecniche è l'edificio della Chase Manhattan Bank a New York (*si veda l'illustrazione in questa pagina*).

Progettato dalla società di progettazione Skidmore, Owings & Merrill e dagli ingegneri costruttori Weiskopf & Pickworth, venne edificato fra il 1957 e il 1961: è stato il primo fabbricato con pareti in vetro alte piú di 240 metri. Esso riassumeva necessariamente la scienza delle costruzioni di tutta un'epoca; ne venne fatta l'analisi delle strutture su modello e fu illustrato in *Structure: An Architect's Approach* di H. Seymour Howard.

Il Chase Manhattan Building è alto 60 piani, divisi in tre luci trasversali e in nove longitudinali. Il regolamento edilizio di New York City prescrive che gli edifici devono essere in grado di resistere a una spinta del vento pari a 97,65 chilogrammi per metro quadrato (20 libbre per piede quadrato) su tutte le superfici situate a oltre 30 metri dal suolo (la pressione frontale del vento su un edificio è proporzionale al quadrato della velocità del vento stesso; la formula empirica usata negli Stati Uniti per mettere in relazione queste grandezze è $P = 0{,}00487V^2$ ove P è la spinta del vento in chilogrammi per metro quadrato e V è la velocità del vento in chilometri all'ora. Una spinta di 97,65 chilogrammi per metro quadrato corrisponde quindi a una velocità del vento di 141,6 chilometri all'ora).

Per il Chase Manhattan Building una spinta del vento di 97,65 chilogrammi per metro quadrato corrisponde a una sollecitazione frontale su ogni luce di 191 420 chilogrammi. La risultante della spinta distribuita dal vento è applicata normalmente alla facciata a un'altezza di 136 metri. Gli sforzi di taglio e di flessione raggiungono il massimo valore al livello del suolo (cioè al piano terra), dove il momento flettente per ciascuna luce è 191 420 × 136, ovvero poco piú di 26 000 000 di chilogrammetri.

L'azione che il vento esercita su un fabbricato a forte sviluppo verticale, edificato in una zona urbana avente un'elevata densità di costruzioni, è piuttosto complessa, ma deve essere prevista dall'ingegnere per poter progettare la sua ossatura in acciaio e dimensionarne gli elementi. Il progetto per la torre di Parigi mostra che Eiffel aveva individuato i tre effetti principali: la flessione dell'intera struttura sotto l'azione della spinta orizzontale del vento, la tendenza della struttura stessa a rovesciarsi ruotando intorno alla linea di base sottovento e la sollecitazione di taglio presente in tutti i montanti. Eiffel si era anche reso conto che i momenti flettenti su una struttura costruita in un materiale elastico come il ferro dovevano inevitabilmente tradursi in uno spostamento misurabile laterale della sua sommità, provocando delle oscillazioni che dovevano essere contenute entro ampiezze tali da non pregiudicarne la stabilità.

Esperimenti condotti nella galleria del vento hanno messo in evidenza ulteriori difficoltà. La vicinanza di altri edifici di grande altezza provoca una turbolenza nelle correnti di aria e inoltre l'effetto Venturi negli stretti spazi delimitati dai fabbricati può aumentare la velocità del vento. Forti spinte del vento accompagnate da improvvise variazioni di direzione inducono flessioni alternate nell'intelaiatura dell'edificio. La formazione e la diffusione ripetuta di vortici, dovuti allo sbarramento delle correnti d'aria, dà luogo a oscillazioni nel piano perpendicolare alla direzione del vento. Infine pressioni non uniformi, piani a sbalzo e l'impiego di montanti situati all'esterno del piano della parete sottopongono la struttura vuoi a torsione vuoi a flessione, e talvolta generano anche pressioni negative, ossia pressioni dell'aria all'interno dell'edificio superiori a quella esterna.

L'ingegnere deve progettare un'intelaiatura e calcolare forme e aree delle sezioni degli elementi strutturali, in modo che la struttura medesima resista all'insieme delle sollecitazioni (almeno per tutta la vita prevista) e non si sposti sotto l'azione del vento in modo sgradevole o tale da turbare gli occupanti. La prima fase della progettazione è la traduzione dei vari effetti del vento in azioni quantitative meccaniche sull'intelaiatura. Il metodo abituale attualmente applicato è un ritorno al metodo di calcolo usato da Eiffel: un edificio di grande altezza è considerato una mensola verticale incastrata nel terreno e sottoposta alle spinte del vento.

Quando una trave (o un edificio) incastrata è sollecitata a flessione, la metà verso il lato piú vicino al carico è convessa mentre l'altra metà è concava. La metà convessa è in tensione con valori di sollecitazione compresi, in teoria, fra zero in corrispondenza dell'asse neutro di mezzeria e un massimo in corrispondenza della fibra esterna. La metà concava è compressa con i medesimi valori di sollecitazione. È possibile fare la stessa generalizzazione per una trave semplice sottoposta a flessione (una trave appoggiata alle estremità), salvo che in questo caso risulta concava la metà verso il lato ove è applicato il carico. La trave è sottoposta anche a sollecitazioni assiali di taglio che sono massime nella zona dove tensione e compressione si avvicinano allo zero, e a sollecitazioni trasversali di taglio che raggiungono il massimo in corrispondenza dell'incastro nel caso della mensola e degli appoggi in quello della trave semplice (*si veda l'illustrazione nella pagina seguente*).

In un fabbricato assimilato a una mensola, supposto soggetto alla sola azione del vento, i montanti sopravento sono in tensione e quelli sottovento in compressione, essendo inoltre presenti sforzi di taglio negli elementi strutturali paralleli alla direzione del vento e in tutti i pilastri. In questi ultimi gli sforzi maggiori sono prodotti dai venti che soffiano sulle facciate piú larghe dell'edificio, perché in tal caso nella profondità dell'edificio vi è un minor numero di montanti sottoposti a taglio e flessione.

Nel caso del Chase Manhattan Building a un vento che soffia in direzione normale a una delle facciate larghe devono resistere quattro pilastri per ciascuna luce. Con una spinta di 97,65 chilogrammi per metro quadrato i pilastri sopravento sono sottoposti a una sollecitazione di tensione di 567 000 chilogrammi e quelli sottovento a una sollecitazione di compressione di 712 000 chilogrammi. I due pilastri interni sono invece sottoposti a carichi molto inferiori. Le sollecitazioni di taglio sono invece distribuite in modo contrario, cioè con il carico maggiore sui pilastri interni.

La rigidità della struttura sotto la azione di questi carichi è assicurata da tre tipi di controventatura. In tutti i piani della luce centrale della sezione trasversale sono state inserite delle travi a K (che in questo caso assumono l'aspetto di lettere K disposte orizzontalmente). Controventi a ginocchio disposti fra solaio e solaio irrigidiscono tutte le luci della sezione longitudinale nelle quali non interferiscono con la distribuzione degli spazi o con i movimenti interni, mentre in tutto il fabbricato sono usate controventature a portale. Nelle luci centrali sia della sezione trasversale sia di quella longitudinale gli elementi diagonali si estendono dalle fondazioni alla copertura. Due sono le ragioni per tale concentrazione di membratura di irrigidimento: le luci centrali, analogamente alle fibre di mezzeria di una trave inflessa, giacciono lungo l'asse secondo il quale la struttura tende con maggiore facilità a dividersi sotto la spinta del vento. Inoltre i pozzi degli ascensori situati in queste luci sono zone di debolezza per complesse e concentrate sollecitazioni che si manifestano lungo i loro bordi.

L'ingegnere deve tener conto anche di altri due effetti del vento. In primo luogo deve assicurarsi che l'edificio non possa essere rovesciato dal vento. Ciò può essere verificato con un bilancio dei momenti: se il fabbricato deve restare in posizione verticale il momento prodotto dal vento che tende a rovesciare la costruzione intorno alla linea di base sottovento deve essere inferiore al momento dovuto al peso proprio che tende a raddrizzarla intorno alla stessa linea. Per il Chase Manhattan Building il momento del vento per luce, secondo quanto è stato detto sopra, è di 26 milioni di chilogrammetri. Il peso proprio di ogni luce pari a 6940 tonnellate è applicato con un braccio di 17,13 metri (distanza fra la mezzeria del fabbricato e la linea di base sottovento) e genera un momento in senso contrario al precedente pari a 118,9 milioni di chilogrammetri. La stabilità dell'edificio sotto la spinta del vento è quindi assicurata, poiché il momento di bilanciamento è quasi cinque volte superiore al momento di rovesciamento.

L'ultima grandezza che deve essere determinata dall'ingegnere è lo spostamento o oscillazione indotta dal vento. Stabilito un valore limite per la deviazione l'intelaiatura viene progettata in modo da evitare deformazioni che superino il limite prefissato. Per gli edifici piú vecchi si imponeva di norma un'oscillazione limite pari a 0,003 h, ove h è l'altezza; per quelli piú recenti che superano 240 metri il limite è stato ridotto a 0,0015 h, valore adottato per il Chase Manhattan Building: per questo fabbricato è consentita quindi un'oscillazione di 366 millimetri, ov-

vero 183 millimetri per parte rispetto all'asse.

Il sistema costruttivo del Chase Manhattan Building potrebbe sembrare l'ultima parola in fatto di controventamento interno, e in un certo senso si poteva considerare esatta tale affermazione. Il numero delle membrature di irrigidimento esistenti del traliccio è immenso: vi sono quasi 3000 elementi diagonali, molte migliaia di piastre e di staffe in ferro e milioni di chiodi. Un peso straordinario e molte ore di lavoro sono state dedicate ad assicurare la rigidità dell'edificio sotto la spinta del vento.

Con la continua richiesta di fabbricati sempre piú alti, qualunque potesse essere il danno che ne derivava all'ambiente urbano, divenne tuttavia evidente che occorreva escogitare dei metodi di controventamento meno costosi. Infatti all'inizio degli anni 1960 parecchi ingegneri svilupparono contemporaneamente un nuovo procedimento.

Il principio fondamentale di questa tecnica è una variazione contemporanea dei concetti su cui si basa il progetto della Torre Eiffel, malgrado che questi non fossero probabilmente noti ai progettisti. I principali innovatori sono stati Fazlur R. Khan e Myron Goldsmith dell'ufficio di Chicago della Skidmore, Owings & Merrill, e numerosi ingegneri facenti parte dello staff di progettisti della ditta Worthington, Skilling, Helle e Jackson di Seattle.

Il ragionamento di questi progettisti era fondato sull'ipotesi che se fosse possibile costruire un edificio che si comportasse come un tubo cavo a pareti solide, un qualcosa di simile a una gigantesca scatola, oppure come un traliccio tubolare irrigidito, analogo alle strutture di supporto dei primi viadotti ferroviari, si otterrebbe una struttura il cui comportamento sarebbe quello di una mensola rigida e le cui pareti esterne sopporterebbero tutte le forze orizzontali, con conseguente drastica riduzione della proliferazione dei controventi interni.

Tale fine può essere raggiunto in vari modi, due dei quali sono illustrati dalle prime opere realizzate secondo tale principio. Una è l'International Business Machines Building a Pittsburgh, eretto, fra il 1922 e il 1963 dagli ingegneri Worthington, Skilling, Helle e Jackson su progetto degli architetti Curtis e Davis. Il fabbricato è una scatola cava rigida con pareti reticolari, una struttura ideata nel 1820 da Ithiel Town, usata molto tempo fa nei ponti, ma mai nei fabbricati. L'altra è rappresentata dai Chestnut-DeWit Apartments a Chicago, edificio costruito contemporaneamente a quello della IBM. È opera di Skidmore, Owings & Merrill, con progetto costruttivo di Khan, ed è costituita da un tubo cavo in cemento armato le cui pareti formano un'ossatura monolitica composta da una fitta schiera di montanti rastremati e da corte travature.

La deflessione di una mensola induce almeno quattro tipi di sollecitazione. Tensione e compressione (*a sinistra*) raggiungono il valore massimo in corrispondenza delle fibre esterne e si avvicinano a zero in corrispondenza dell'asse neutro di mezzeria. Nel lato convesso si ha tensione (*in colore*), e in quello concavo compressione (*in grigio*). Sono presenti anche sollecitazioni di taglio (*a destra*). Il taglio longitudinale (*in colore*) è provocato dal contrasto fra sforzi di tensione e di compressione e raggiunge il massimo valore in mezzeria; il taglio trasversale (*in grigio*) è massimo all'incastro della struttura. In un edificio considerato come una mensola sottoposta alla spinta del vento possono essere presenti anche forze che tendono a rovesciare e a torcere la struttura.

L'edificio piú spettacolare e complesso realizzato applicando la tecnica citata è il John Hancock Center a Chicago alto 100 piani, opera di Skidmore, Owings & Merrill, eretto su progetto costruttivo di Khan fra il 1965 e il 1970 (*si veda l'illustrazione a pagina 87*). Si presenta come una mensola tubolare irrigidita a profilo rastremato ed è l'equivalente quasi esatto di un'antica torre di sostegno a traliccio di ponti in ferro sotto forma di grattacielo. Ciascuna delle quattro facciate è costituita da pilastri e da travi rigidamente vincolate fra loro mediante immense membrature diagonali incrociate che formano delle travature a X larghe cinque pannelli e alte 18 piani e da travi a K nel pannello terminale alto 10 piani. Pilastro, trave e diagonale si incontrano in un punto dove sono rigidamente connessi con una grossa piastra di rinforzo.

Il risultato costruttivo di questo « supertraliccio » impeccabilmente progettato è che tutti gli sforzi di flessione prodotti dal vento e tutti i carichi assiali nei pilastri determinati da sollecitazioni di tensione e di compressione dovute al vento sono eliminati dalla struttura interna e trasferiti sulle pareti a traliccio. Anche le diagonali contribuiscono a distribuire le forze, cosicché qualsiasi configurazione di carichi verticali su un montante a una certa quota è uniformemente ripartito fra tutti i montanti alla base. Ancora piú notevole, perché contrario a quanto si verifica normalmente in un traliccio, è il fatto che le diagonali sono quasi sempre compresse, indipendentemente dall'azione del vento e per questo motivo sono state definite « pilastri inclinati ». Le conseguenze economiche del progetto sono eccezionali quanto l'aspetto della costruzione: la struttura contiene solo 145 chilogrammi di acciaio per ogni metro quadrato di superficie in pianta, mentre negli edifici a traliccio convenzionale si hanno di norma valori compresi fra 195 e 245 chilogrammi per metro quadrato.

Esistono anche altri metodi costruttivi per le mensole tubolari rigide. La struttura muraria che si avvicina con la migliore approssimazione a un diaframma piano è quella dello Standard Oil Building di Chicago. Progettata da Edward Durell Stone e dalla Società Perkins & Will e iniziata nel 1970, la torre ha 80 piani raggiungendo un'altezza di 414,5 metri. Le facciate, dotate di finestre piccole e notevolmente distanziate, sono chiuse da piastre di acciaio piene irrigidite da prismi triangolari a forma di sperone che, oltre alla funzione strutturale di contrafforti, alloggiano condutture, condotti e tubature.

Altra variazione è il World Trade Center di New York, costituito da due torri di 110 piani alte 411,5 metri. Il Centro, iniziato nel 1968, è stato progettato da Minoru Yamasaky, Emery Roth & Sons e da Worthington, Skilling, Helle e Jackson. Il fitto disegno esterno delle torri è realizzato con gli elementi verticali molto ravvicinati di tralicci Vierendeel. Si tratta di tralicci privi di membrature diagonali, la cui rigidità è assicurata da giunti progettati in modo da resistere alla flessione e al taglio. Un sistema analogo è stato adottato nell'ossatura delle pareti del Sears, Roebuck and Company Building di Chicago, iniziato nel 1970, che con i suoi 442 metri è attualmente il piú alto edificio del mondo. Ne sono stati architetti e costruttori Skidmore, Owings & Merrill.

Un terzo sistema è stato impiegato nel Boston Company Building, edificato fra il 1969 e il 1970 a Boston. Ne sono stati architetti Pietro Belluschi e Emery Roth, mentre i costruttori appartenevano allo staff di James Ruderman. I tralicci a K disposti nell'intelaiatura delle pareti, cui si aggiunge una struttura ad anello al livello del secondo piano, trasmettono tutte le spinte del vento e la maggior parte dei carichi gravitazionali a quattro poderosi pilastri angolari rastremati con sezione a croce. Questo sistema si è finora dimostrato il piú economico: la quantità di acciaio è di appena 102 chilogrammi per metro quadrato di superficie in pianta.

Tutti gli antenati di questi sistemi costruttivi si ritrovano nell'evoluzione dei ponti a travi reticolari nel corso degli ultimi 150 anni; gli edifici « tubolari » possono essere considerati ponti a sbalzo vincolati a una estremità. Al momento in cui verranno introdotti sistemi costruttivi analoghi a quelli dei ponti sospesi, come in effetti si sta già verificando, l'analogia fra il progetto dei ponti in acciaio e quello dei grattacieli a traliccio sarà pressoché completa.

Un traliccio esterno realizzato con immense membrature in acciaio sopporta tutte le spinte del vento nell'edificio alto 100 piani del John Hancock Center di Chicago. In ogni facciata le controventature formano delle travature a X larghe cinque pannelli e alte 18 piani e delle travi a K negli ultimi 10 piani. La torre è una mensola tubolare irrigidita ed è l'equivalente in forma di grattacielo delle torri di sostegno dei ponti in ferro del XIX secolo. Gli sforzi di flessione indotti dal vento sono eliminati dalla struttura interna e trasferiti sulle pareti a traliccio. Inoltre qualsiasi configurazione di carichi verticali è uniformemente ripartita fra i montanti. Architetti e costruttori appartenevano all'ufficio di Chicago della società Skidmore, Owings & Merrill.

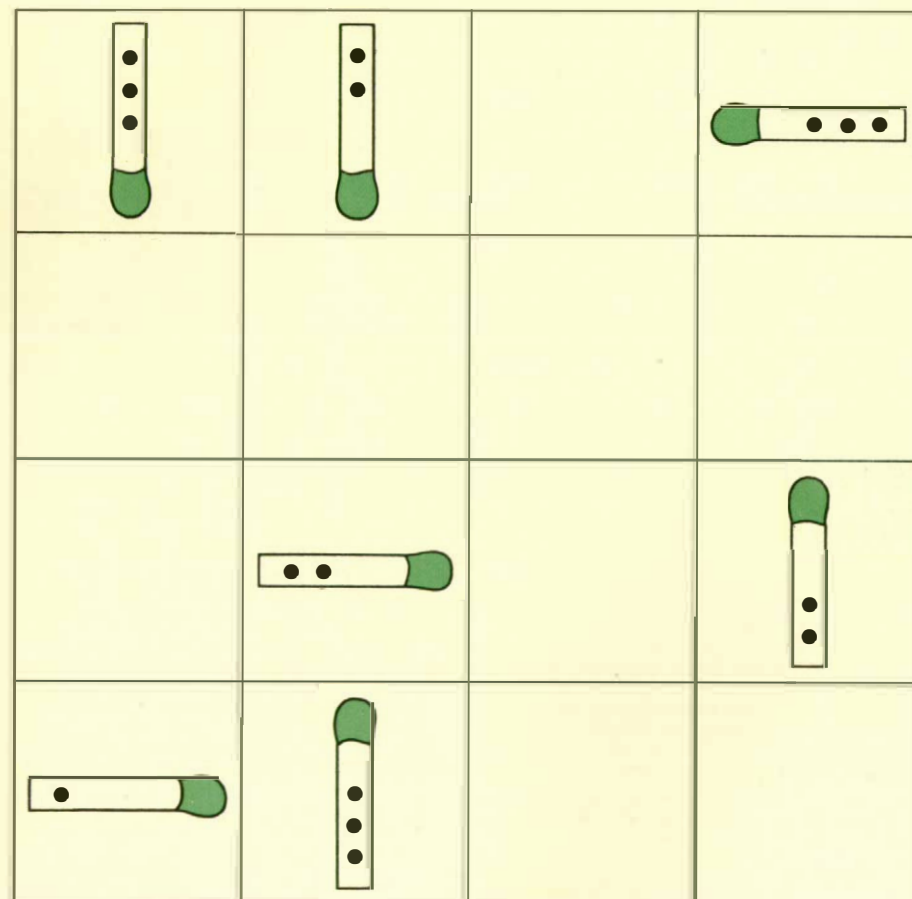
GIOCHI MATEMATICI

di Martin Gardner

Nuova raccolta di problemi di vario genere

Di tanto in tanto presento una raccolta di problemi di vario genere che dovrebbero esser nuovi per la maggior parte dei lettori e che non richiedono conoscenze di matematica superiore. Le soluzioni di tutti questi problemi saranno pubblicate il mese prossimo.

1. Disegnate una grande scacchiera di 4 caselle per 4 su un foglio di carta, procuratevi 16 fiammiferi svedesi e siete pronti a mettervi all'opera con questo nuovo tipo di solitario. I fiammiferi rappresentano frecce orientate nella direzione della capocchia. Disegnate un puntino su entrambi i lati di un fiammifero, due puntini su entrambi i lati di 8 fiammiferi e tre puntini su entrambi i lati di 7 fiammiferi. Quando un fiammifero viene collocato su una casella della scacchiera orientato verso nord, sud, est o ovest, il punto singolo significa che esso punta alla casella immediatamente contigua, due punti significano che esso punta alla seconda casella e tre alla terza.

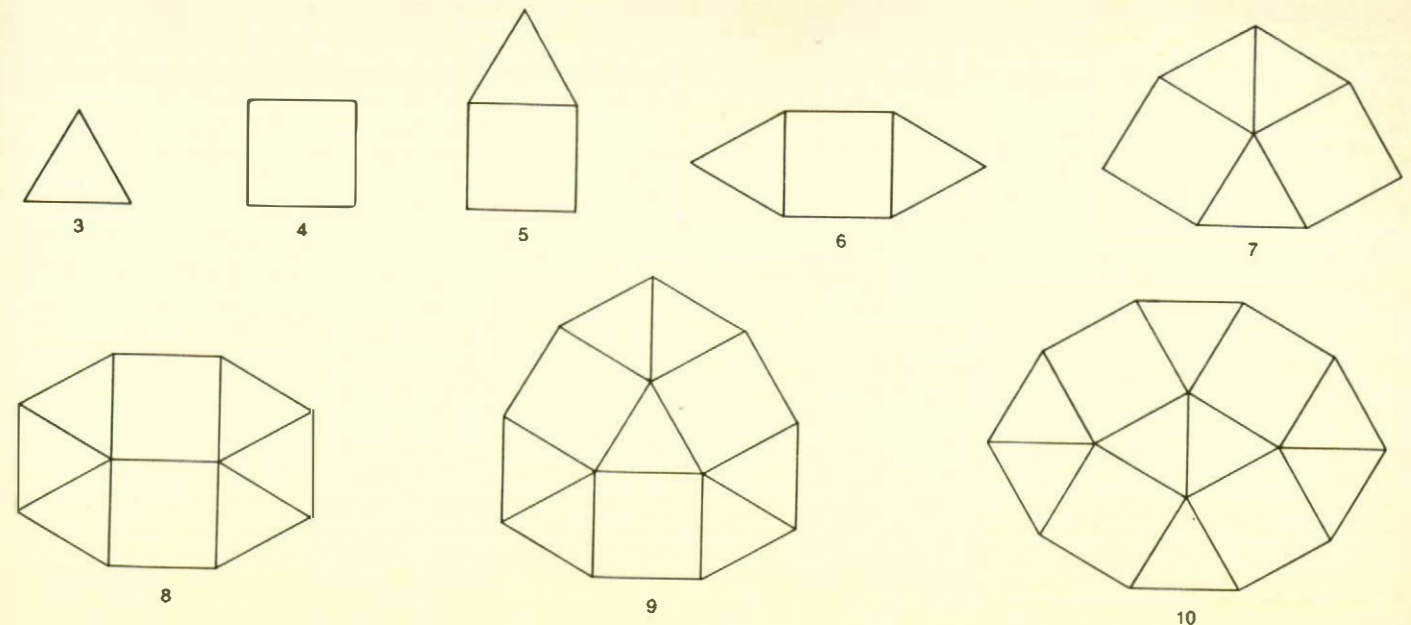
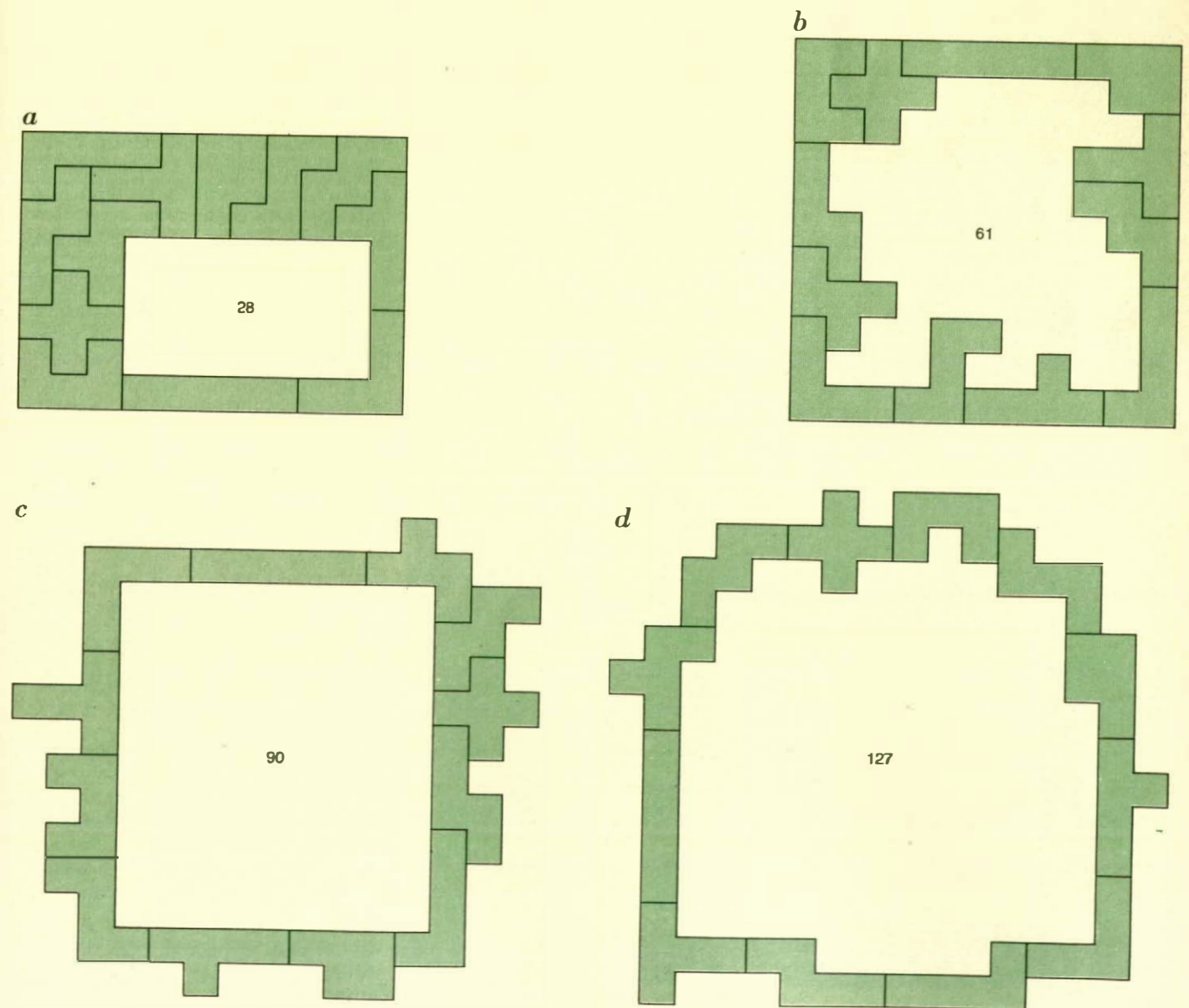
Sette fiammiferi possono essere collocati in modo da configurare un giro chiuso (*si veda la figura in basso*). Prendendo l'avvio da uno qualsiasi dei sette fiammiferi, collocate il dito indice sulla casella da esso indicata. La freccia che si trova in tale casella vi darà l'indicazione per la mossa successiva. Seguite le frecce fino a tornare (in sette mosse) al punto di partenza. Il problema consiste nel collocare tutti e sedici i fiammiferi, uno in ciascuna casella, in modo che essi configurino un giro chiuso che visiti ogni casella. Esistono soltanto due soluzioni, non contando rotazioni e riflessioni.

Un giro di frecce chiuso.

Il giro avrà una lunghezza di $1 + (2 \times 8) + (3 \times 7) = 38$. Non è difficile dimostrare che si tratta del giro chiuso piú lungo che possa essere fatto sulla scacchiera usando ogni combinazione dei tre tipi di frecce. Brian R. Barwell, un ingegnere britannico che introdusse il problema nel « Journal of Recreational Mathematics » (ottobre, 1969) trovò che è possibile soltanto un altro giro di lunghezza massima. Esso richiede 6 frecce con 3 puntini, 10 frecce con 2 puntini e nessuna freccia con 1 puntino. I lettori sono invitati a cercare tutti e tre questi percorsi.

Le frecce sono naturalmente solo un modo conveniente per configurare un giro chiuso di lunghezza massima, percorso da una torre degli scacchi che si fermi esattamente una volta in ogni casella. (Giri di questo tipo compiuti dalla regina sono meno interessanti perché sono molto numerosi, i giri compiuti dagli alfieri non possono chiudersi e non possono visitare tutte le caselle e i giri di cavallo non possono variare in lunghezza.) La scacchiera 2×2 è banale e quella 3×3 viene analizzata facilmente. (Il suo giro massimo ha una lunghezza di 14 caselle.) A quanto mi consta, le scacchiere 5×5 e superiori non sono state ancora studiate.

2. Recentemente mia moglie e io intervenimmo a un party a cui erano presenti altre quattro coppie di coniugi. Ebbero luogo varie strette di mano. Nessuno strinse la mano a se stesso (o a se stessa), nessuno strinse la mano al proprio coniuge e nessuno strinse la mano piú di una volta alla stessa persona.

Dopo che tutte le strette di mano ebbero avuto luogo, chiesi a ciascuno, compresa mia moglie, quante mani avesse stretto. Con mia sorpresa, ciascuno diede una risposta diversa. Quante mani strinse mia moglie? (Da Lars Bertil Owe di Lund, Svezia).

3. Supponiamo che sia disponibile un numero illimitato di quadrati e triangoli equilateri di cartone, ciascuno con lati unitari. Usando questi elementi è facile formare poligoni convessi aventi un numero di lati compreso fra 3 e 10 (*si veda la figura in alto nella*

Poligoni convessi con un numero di lati compreso fra 3 e 10.

Problemi di recinzione con pentamini.

pagina precedente). Siete in grado, utilizzando gli stessi elementi, di formare un poligono convesso di 11 lati? E qual è il numero massimo di lati che un poligono convesso formato da tali elementi può avere? (Entrambe le domande si fondano sulla soluzione di Michael Goldberg, in « The American Mathematical Monthly », maggio 1969, a un problema posto l'anno prima da Joseph Malkewitch.)

4. Si valuti, rispetto alla sua veridicità o falsità, ciascuna delle 10 proposizioni seguenti. (Da David L. Silverman di West Los Angeles.)

1) Esattamente una proposizione in quest'elenco è falsa.

2) Esattamente due proposizioni in quest'elenco sono false.

3) Esattamente tre proposizioni in quest'elenco sono false.

4) Esattamente quattro proposizioni in quest'elenco sono false.

5) Esattamente cinque proposizioni in quest'elenco sono false.

6) Esattamente sei proposizioni in quest'elenco sono false.

7) Esattamente sette proposizioni in quest'elenco sono false.

8) Esattamente otto proposizioni in quest'elenco sono false.

9) Esattamente nove proposizioni in quest'elenco sono false.

10) Esattamente dieci proposizioni in quest'elenco sono false.

Foglio di carta piegato con un motivo a rete di pollaio.

5. Victor G. Feser, della Saint Louis University, ha proposto quattro problemi concernenti l'area massima, da risolvere ciascuno usando l'intera serie dei 12 pentamini. Tre sono stati risolti, e anche il quarto probabilmente lo è.

1) Si formi un « recinto » rettangolare attorno alla massima superficie rettangolare. È stato dimostrato che la massima superficie recintata è di 4 × 7 (*si veda la soluzione « a » nella figura alla pagina precedente in basso*).

2) Si formi un recinto rettangolare attorno alla massima superficie, di qualsiasi forma. Il massimo è 61 quadrati unitari (*si veda la soluzione « b » nella figura*).

3) Si formi un recinto di forma qualsiasi attorno alla massima superficie rettangolare. Il 9 per 10 è il massimo (*si veda « c » nella figura*).

4) Si formi un recinto, di qualsiasi forma, attorno alla massima superficie di qualsiasi forma. (Come nei problemi precedenti, il recinto deve avere in tutti i punti almeno uno spessore unitario.) Chris Lindstedt, di Göteborg, in Svezia, Wade E. Philpott di Lima, Ohio, e io trovammo soluzioni di 127 quadrati, come quello presentato in *d* nella figura. Si credeva che questo fosse il massimo, finché Donald E. Knuth, lo studioso dei calcolatori, della Stanford University, li portò recentemente a 128. Knuth possiede una dimostrazione non formale del fatto che il numero 128 non può essere superato.

I lettori troveranno piacevole e difficile trovare una soluzione che ammetta 128 quadratini. Se qualcuno potrà superare questo numero o dimostrare con un calcolatore che si tratta del valore massimo, gliene sarò molto grato. Per informazioni sui problemi primo e terzo si veda l'articolo di Feser, *Pentomino Farms* (Recinti di pentamini), nel « Journal of Recreational Mathematics », gennaio 1968, pagine 55-61, e le discussioni nei numeri dell'ottobre 1968, pagina 243, e del luglio 1969, pagina 187.

6. Una cucina ha un pavimento non ben livellato. Non è che ci siano mattonelle sporgenti, ma il carattere casualmente ondulatorio del linoleum è tale che quando si cerca di sistemare un piccolo tavolo quadrato di quattro gambe, una gamba di solito non poggia, facendolo vacillare. Se non si fa caso al fatto che il piano del tavolo rimanga inclinato, è sempre possibile trovare un punto in cui le quattro gambe del tavolo siano saldamente poggiate sul pavimento? Oppure un pavimento può essere ondulato in modo tale che non sia disponibile alcuna posizione del genere? Il problema può essere risolto con una dimostrazione semplice ed elegante. (Da Miodrag Novakovic, di Belgrado, e Ken Austin di Chesham, in Inghilterra.)

7. Questo strano gioco da salotto proviene da Tan Hock Chuan, un mago di professione che vive a Singapore. Egli lo descrisse in una lettera a Johnnie Murray, un prestigiatore dilettante di Portland, nel Maine, che me lo passò.

Un foglio di carta bianco di cm 20 × 12 (la metà di un foglio per dattilografia va benissimo) viene siglato da uno spettatore, in modo da poter essere piú tardi identificato. Il mago tiene il foglio dietro le spalle (o sotto un tavolo) per circa 30 secondi; quando torna a farlo vedere, esso è ricoperto da pieghe che formano un motivo a mosaico costituito da un disegno esagonale regolare (*si veda la figura della pagina a fronte*). In che modo avviene ciò? Chi esegue il gioco viene accusato di solito di aver compresso il foglio contro un pezzo di rete metallica, come quella usata per recingere i pollai, ma le pieghe sono eseguite in realtà usando soltanto le mani.

8. Alcuni tra i piú bei problemi di scacchi moderni sono quelli che richiedono un'analisi retrograda: ossia la deduzione delle mosse che devono aver preceduto la posizione data.

Il problema presentato nella figura di questa pagina è un'eccellente introduzione a questo ambito. Esso è incantevolmente semplice e a prima vista apparentemente senza soluzione. Il problema mi fu inviato due anni fa da Robert Rosenwald di New Hope, in Pennsylvania, il quale non ne conosceva l'origine. La figura illustra una posizione in una convenzionale partita di scacchi subito dopo che il re bianco è stato accidentalmente fatto cadere oltre il bordo della scacchiera. Dove si trovava il re bianco e qual è stata l'ultima mossa del bianco?

Dove si trovava il re bianco?

9. Per convenzione, il valore di una scala esponenziale come

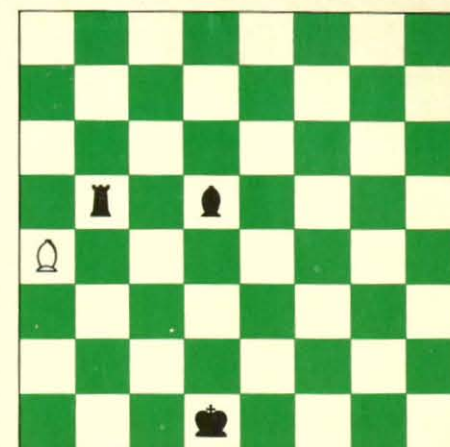
$$2^{2^{2^{2^{2}}}}$$

è calcolato partendo dall'esponente piú alto e scendendo verso il basso. I due esponenti piú in alto danno 4, dopo di che si ha $2^4 = 16$ e $2^{16} = 65\,536$. Quanto fa $2^{65\,536}$? Alcuni anni fa, Geoffrey W. Hoffmann della Germania Federale mi inviò la soluzione, ottenuta con un calcolatore. È un numero che comincia con 20035... e ha 19 729 cifre. Se si aggiunge un altro esponente 2, si avrà un numero che non sarà mai calcolato poiché la risposta, come dice Hoffmann, richiederebbe l'età dell'universo per essere calcolata dal calcolatore e lo spazio dell'universo per essere stampata. Anche una scala esponenziale brevissima, comprendente tre soli esponenti 9, dà $9^{387\,420\,489}$, un numero comprendente piú di 360 milioni di cifre. Nel 1933 E. Skewes pubblicò un articolo in cui dimostrò che se $\pi(x)$ è il numero di numeri primi inferiore a x, e $\mathrm{li}(x)$ è la funzione logaritmica integrale, allora $\pi(x) - \mathrm{li}(x)$ è positivo per taluni valori di x inferiori a

$$10^{10^{10^{34}}}$$

un numero intero di cui si dice che è il piú grande numero noto a svolgere una funzione in un teorema non banale.

Nel 1971 Aristid V. Grosse, uno tra i primi fisici atomici alla Columbia University nel 1940 (ora presidente dei Germantown Laboratories, Inc., affiliati al Franklin Institute), iniziò uno studio di scale esponenziali di numeri identici che sono calcolati nella direzione opposta (verso l'alto) e delle loro relazioni a scale discendenti. Per scale di entrambi i tipi egli coniò il termine « polipotenze. » Scale esponenziali composte da due x vengono chiamate « dipotenze », di tre x « tripotenze » e cosí via secondo i prefissi greci. Il valore di x può essere razionale o irrazionale, trascendente, complesso o infine immaginario. Nella maggior parte dei casi le polipotenze presentano un valore singolo, sono continue e differenziabili. Poiché in ogni polipotenza 1 elevato a 1 è uguale a 1, tutte queste funzioni e le loro derivate, se collocate in un grafico in funzione di x, si incrociano in corrispondenza di $x = 1$, e i loro valori allo 0 sono i limiti quando x si approssima a 0. Le osservazioni di Grosse, che riempiono già molti volumi, conducono a un intrico lussureggiante di teoremi insoliti oltre che a nuove classi di numeri.

Le dipotenze sono ovviamente identiche sia che vengano calcolate verso l'alto o verso il basso, mentre per tutte le altre polipotenze superiori le due direzioni danno numeri diversi. Il tripletto del 9, per esempio, se viene calcolato verso l'alto, è un numero di sole 77 cifre. Il numero di modi esistenti per « sviluppare » una scala esponenziale (per mezzo di una diversa collocazione delle parentesi) è 2^{n-2}, dove n è il numero delle x. In tutti i casi in cui il calcolo si esegue verso l'alto si ottiene il valore minimo, mentre il procedimento inverso dà il valore massimo. In quanto segue, le frecce indicano questi valori massimi e minimi.

Che cosa accade quando si stabilisce un'eguaglianza fra scale di diversa lunghezza calcolate nei due versi opposti? Se un tripletto di x calcolato verso l'alto è uguale a un tripletto di x calcolato verso il basso, $x = 2$. (Escludiamo il caso di $x = 1$, in quanto banale.) Ogni x in piú nella scala ascendente aumenta di 1 il valore della x. Se tre x calcolate verso il basso danno un valore uguale a quello di quattro x calcolate verso l'alto, $x = 3$; se tre x calcolate verso il basso sono uguali a cinque x calcolate verso l'alto, $x = 4$ e cosí via.

Come introduzione alle polipotenze, chiediamo al lettore di risolvere le tre equazioni qui sotto, le quali cominciano una serie con tetrapotenze discendenti a sinistra:

$$\overleftarrow{x^{x^{x^{x}}}} = \overrightarrow{x^{x^{x}}}$$

$$\overleftarrow{x^{x^{x^{x}}}} = \overrightarrow{x^{x^{x^{x}}}}$$

$$\overleftarrow{x^{x^{x^{x}}}} = \overrightarrow{x^{x^{x^{x^{x}}}}}$$

I lettori possono divertirsi a investigare su scale di x frazionarie, di reciproci di x e forme piú strane. Grosse ha sviluppato anche il concetto di una polipotenza perfetta, che è quella per la quale x sta alla x-esima potenza (ascendente o discendente) un numero x di volte. (Esempio: π alla π-esima potenza, π volte in senso ascendente, è uguale a 588 916,326+.) L'operazione inversa alle polipotenze viene denominata da Grosse « poliradice ». Questi campi sono mai stati investigati prima? Nonostante sforzi considerevoli, né lui né io abbiamo trovato riferimenti precisi. Entrambi saremmo grati a chi fosse in grado di darci qualche segnalazione in proposito.